Cent. Eur. J. Med. • 4(1) • 2009 • 37-48
DOI: 10.2478/s11536-008-0085-z

VERSITA

Central European Journal of **Medicine**

Determination and Modelling of Clinical Laboratory Data of Healthy Individuals and Patients with End-Stage Renal Failure

Research Article

Agelos Papaioannou[1*], George Karamanis[2], Ioannis Rigas[3], Thomas Spanos[4], Zoe Roupa[5]

[1] *Section of Clinical Chemistry and Biochemistry, Department of Medical Laboratories, Faculty of Health and Care, Technological and Education Institute of Larissa, 41110 Larissa, Greece*

[2] *Biochemical Laboratory, General Hospital of Kavala, 65201 Kavala, Greece*

[3] *Department of Animal Production, Technological and Education Institute of Larissa, 41110 Larissa, Greece*

[4] *Education and Technological Institute of Kavala, 65403 Kavala, Greece*

[5] *Nursing Department, Technological and Education Institute of Larissa, 41110 Larissa, Greece*

Received 9 June 2008; Accepted 21 October 2008

Abstract: The analyses of 18 biochemical parameters (alanine aminotransferase, albumin, aspartate aminotransferase, calcium, cholesterol, chloride, creatinine, iron, glucose, γ- glutamyl transferase, alkaline phosphatase, phosphorus, potassium, sodium, total protein, triglycerides, uric acid, and urea nitrogen) were performed for 166 healthy individuals and 108 patients with end-stage renal failure (ESRF). The application of cluster analysis proved that there were points of similarity among all 18 biochemical parameters that formed major groups; these groups corresponded to the authors' assumption of the existence of several overall patterns of biochemical parameters that may be termed "enzyme-specific"; "general health indicator"; "major component excretion"; "blood-specific indicator"; and "protein-specific". These patterns also appear in the subsets of males and females that were obtained by separation of the general dataset. In addition, the performance of factor analysis similarly proved the validity of this assumption. This projection and modelling method indicated the existence of seven latent factors, which explained 70.05% of the total variance in the system for healthy individuals and more than 72% of the total variance in the system for patients with ESRF. All these results support the probability that a general health indicator could be constructed by taking into account the existing classification groups in the list of biochemical parameters.

Keywords: *Healthy individuals • End-stage renal failure patients • Biochemical parameters • Cluster analysis • Factor analysis*

1. Introduction

Chronic renal failure (CRF) is a slowly progressive loss of renal function over a period of months or years and is defined as an abnormally low glomerular filtration rate, which is usually determined indirectly by the creatinine level in blood serum. CRF that leads to severe illness and requires some form of renal replacement therapy (such as dialysis) is called end-stage renal disease (ESRD) [1-9].

ESRD is a complete or nearly complete failure of the kidneys to excrete wastes, concentrate urine, and regulate electrolytes. ESRD occurs when the kidneys are no longer able to function at a level that is necessary for daily life. It usually occurs when chronic renal failure

* *E-mail: papaioannou@teilar.gr*

results in kidney function that is less than 10% of normal. At that point, complications are multiple and severe, and, without dialysis or kidney transplantation, death will occur from bodily accumulation of fluids and waste products [10].

Many diseases can irreversibly damage or injure the kidneys. Acute kidney failure can become chronic if function is not recovered after treatment. Therefore, any process that can cause acute kidney failure can cause chronic kidney failure. The most common causes of CRF in North America and Europe are diabetic nephropathy, hypertension, and glomerulonephritis. Together, these cause approximately 75% of all cases in adults. Certain geographic areas have a high incidence of human immunodeficiency virus (HIV) nephropathy [11-15].

The patient with ESRD usually has a long history of chronic renal failure that has progressed and may have required dialysis to control. The urine volume may decrease, or urine production may stop totally. There is often evidence of complications. Creatinine and urea levels are chronically high, and creatinine clearance is very low. Blood tests are essential for diagnosis and monitoring chronic renal failure, because the results show an increase in levels of urea and creatinine, metabolic waste products that are normally filtered out by the kidneys. The level of potassium in the blood is normal or only slightly increased, but it can become dangerously high when kidney failure reaches an advanced stage or if a large amount of potassium is ingested. The level of triglycerides in the blood is likely to be elevated; the calcium level decreases, and the phosphorus level increases. Dialysis is the usual option for ongoing treatment and is often used while waiting for a suitable transplant opportunity. It is not as efficient as a human kidney, so those with chronic kidney failure usually need to restrict their intake of fluid and of certain foods. They also require additional medication such as iron supplements, phosphate binders, and antihypertensive drugs. Despite the advent of dialysis, most people with advanced kidney failure die within 5 to 10 years [16-21].

The prognosis for patients with chronic renal disease is guarded, because epidemiologic data has shown that all-cause mortality (the overall death rate) increases as kidney function decreases. The leading cause of death in patients with chronic renal disease is cardiovascular disease, regardless of whether there is progression to ESRD [22-25].

The aim of this study is to offer a simple multivariate statistical strategy, in addition to the recommended monitoring procedures, for interpretation and modelling of the laboratory data usually determined in clinical and biochemical laboratories.

2. Material and Methods

We studied the distribution patterns of some analytes commonly assessed in clinical chemistry - biochemistry laboratories for healthy individuals and for patients from Greece with ESRF.

The way in which healthy individuals were selected is described in detail elsewhere [26]. Patients with end-stage renal failure undergoing hemodialysis were being treated at the General Hospital of Kavala. A total of 166 healthy individuals from the Prefectures of Drama and Kavala (females N=94, males N=72), aged 18 to 30 years, and 108 patients with ESRF from the General Hospital of Kavala (females N=45, males N=63), aged 35 to85 years, were tested. The data in this study were derived from the blood samples taken in the biochemical laboratories of the General Hospital of Drama and Kavala (Greece). During each sampling period, blood specimens were collected between 07:30 and 10:00 hours. The subjects prepared by eating a light supper the night before the tests, forgoing alcoholic beverages, fasting overnight, consuming only water for breakfast the day of the tests, and avoiding vigorous exercise. The blood samples were collected in Vacutainer® tubes (Becton Dickinson Co., Rutherford, NJ) free of anticoagulant, according to international specifications [27,28]. The Vacutainer tubes were left for a period of time at ambient temperature in order for the blood clot. The blood serum was separated by centrifugation at 1000 g for 20 minutes, and then the Olympus AU640 analyzer (general hospital of Drama) and Dimension RXL (General Hospital of Kavala) determined the selected concentrations for the biochemical parameters, within a 2-hour period.

The analyses of 18 biochemical parameters [(alanine aminotransferase (ALT), albumin (ALB), aspartate aminotransferase (AST), calcium (Ca), cholesterol (CHOL), chloride (Cl), creatinine (CREA), iron (Fe), glucose (GLU), γ- glutamyl transferase (GGT), alkaline phosphatase (ALP), phosphorus (P), potassium (K), sodium (Na), total protein (TP), triglycerides (TG), uric acid (UA), and urea nitrogen (UREA)] were performed on an Olympus AU640 analyzer (Olympus, Japan) and Dimension RXL at 37°C immediately after centrifugation, according to the methods listed in Table 1.

Before each determination, calibration and internal control of analyzers with calibrators and quality controls proceeded according to the manufacturers' instructions and international literature [30,31]. The reagents provided in commercial kits were used in the analyzer, and the methods were adapted according to the manufacturers' instructions. The water, free from metal

Table 1. Methods used for the determination of the different quantities (37°C).

Variable	Olympus method	Dimension method
Alanine aminotransferase ALT	IFCC w/o p-5'-p	IFCC with (P-5-P)
Albumin ALB	Bromocresol green	(BCP) purple
Aspartate aminotransferase AST	IFCC w/o p-5'-p	IFCC with (P-5-P)
Alkaline phosphatase ALP	IFCC/AMP Buffer	AMP buffer
Calcium Ca	o-cresolphtalein-complex	o-cresolphtalein-complex
Cholesterol CHOL	CHOD/PAP	CHOD/PAP ή CHOD/POD
Chloride Cl	ISE indirect	IMT Indirect
Creatinine CREA	Jaffe´	Jaffe´
Glucose GLU	Hexokinase	Hexokinase (HK/G-6-PDH)
γ-glutamyl transferase GGT	SZASZ	IFCC (adapted)
Iron Fe	TPTZ	Ferene
Phosphorus P	Phosphomolybdate	Phosphomolybdate U.V.
Potassium K	ISE indirect	IMT Indirect
Sodium Na	ISE indirect	IMT Indirect
Total Proteins	Biuret	Biuret
Triglycerides TG	GRO/PAP	(CHOD/PAP or CHOD/POD)
Uric acid UA	Uricase/PAP	Uricase/PAP or Uricase/POD
Urea UREA	Urease U.V.	Urease/GLDH U.V.

ions, had a maximum receptivity of 18.2 Mohm cm at 25° C. Accuracy was checked (and achieved) by two external quality control programs (Radox (RIQAS) and Greece (ESEAP)).

The calculations were performed with the software package STATISTICA 7.0 for Windows.

2.1. Statistical Analysis

Basic statistical and correlation calculations were carried out to provide initial information about the biochemical laboratory data. To evaluate the correlations between the levels of biomarkers of each group, the Pearson correlation coefficients were calculated.

To identify variables independently associated with outcomes, cluster analyses and factor analyses were used. Cluster analysis and factor analysis are multivariate statistical techniques that can be used to interpret biochemical data and assist in clinical laboratory data monitoring and planning [32-41].

Cluster analysis 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 cluster analysis 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.

Cluster analysis was conducted to group biochemical data (1) of healthy individuals and (2) of patients with ESRF by the complete linkage method with squared Euclidean distance measure. This type of analysis was used to link variables in the configuration of a tree with different branches; branches that have linkages closer to each other indicate a stronger relationship among variables or a cluster of variables. The dendrogram generated from tree clustering provides a useful graphic tool for determining the number of clusters that describe underlying processes that lead to spatial variation.

Factor analysis is used to understand the correlation structure of collected data and identify the most important factors contributing to the data structure. In factor analysis, the relationship among a number of observed quantitative variables is represented in terms of a few underlying, independent variables called factors, which may not be directly measured or even measurable. Factor analysis is also used to find associations between parameters, so that the number of measured parameters can be reduced. Known associations are then used to predict unmeasured biochemical quality parameters.

Although not commonly used in laboratory data analysis, several studies have employed factor analysis to interpret and to model the clinical laboratory data [32,34,38,40]. The initial step was the determination of the parameter correlation matrix, which was used to account for the degree of mutually shared variability between individual pairs of biochemical parameters. The second step was the estimation of the eigenvalues and factor loadings for the correlation matrix. Each eigenvalue corresponded to an eigenfactor that

Table 2. Basic statistics for the tested biochemical parameters for male and female healthy individuals (mean value, minimum and maximum values within a certain variable, standard deviation of the mean).

Male healthy individuals					Female healthy individuals				
Variable	Mean	Min	Max	SD	Variable	Mean	Min	Max	SD
GLU	98	82	133	10	GLU	93	74	117	8
UREA	28.8	17.0	40.0	5.8	UREA	25.3	14.0	44.0	6.1
CREA	1.05	0.69	1.26	0.12	CREA	0.89	0.59	1.16	0.10
UA	5.4	3.3	8.0	1.0	UA	3.8	2.2	7.2	0.8
CHOL	161	106	308	37	CHOL	165	65	253	33
TG	94	27	331	51	TG	76	31	207	34
AST	24.0	14.0	54.0	7.8	AST	20.3	13.0	52.0	6.5
ALT	26.0	11.0	74.0	13.3	ALT	18.7	9.0	86.0	13.2
ALP	78.8	34.0	186.0	26.8	ALP	56.1	33.0	97.0	14.9
GGT	20.1	7.0	75.0	10.9	GGT	12.3	5.0	46.0	6.3
K	4.4	3.8	5.2	0.3	K	4.4	3.8	5.2	0.4
Na	141.1	136.0	147.0	2.2	Na	139.3	132.0	144.0	2.4
Cl	103.1	84.0	108.0	3.6	Cl	103.5	99.0	108.0	2.1
Ca	9.8	8.8	10.8	0.5	Ca	9.6	8.5	10.8	0.5
Fe	98.9	41.0	184.0	38.3	Fe	85.9	40.0	180.0	36.4
P	4.0	2.6	5.5	0.7	P	3.9	2.5	5.0	0.5
TP	7.69	6.30	8.50	0.37	TP	7.62	6.50	8.60	0.45
ALB	5.11	3.80	5.90	0.35	ALB	4.90	4.00	6.00	0.40

identified the groups of variables that were most highly correlated among them. The first eigenfactor accounted for the greatest variation among the observed variables, while each subsequent eigenfactor was orthogonal to all preceding factors and provided incrementally smaller contributions to the overall descriptive ability of the model. Because lower eigenvalues might contribute little to the explanatory capability of the data, only the first few factors were needed to account for much of the parameter variability.

In this study, the factor extraction was performed with the method of principal components. The most widely used method for determining how many factors to keep and how many to ignore is the Kaiser criterion, which retains only those factors with eigenvalues >1. This means that each retained factor provides as much explanatory capability as one original variable.

Once the correlation matrix and eigenvalues were obtained, factor loadings were used to measure the correlation between variables and factors. Factor rotation was used to facilitate interpretation by providing simpler factor structure. The factors were rotated so that the observed axes were aligned with a dominant set of variables, which assisted in the understanding of how factors were related to the observed variables. In this study the varimax rotation, a standard rotation method, was used.

3. Results

Basic statistical data (mean value, minimum and maximum values, standard deviation) for the biochemical parameters of the healthy individuals for both the male and female categories are presented in Table 2.

The correlation between the different biochemical test parameters of all 166 healthy individuals showed that the overall significance of many was statistically sound, according to the Pearson test. For several parameters such as CREA/ALB (0.475, $p<0.001$), AST/ALT (0.756, $p<0.001$), CREA/UA (0.624, $p<0.001$), UA/GGT (0.442, $p<0.001$), UA/ALP (0.420, $p<0.001$), ALT/GGT (0.611, $p<0.001$), and TP/ALB (0.743, $p<0.001$), a real logical interpretation ($r > 0.4$) for significance could be offered.

Similarly, basic statistics for the biochemical parameters of the patients with ESRF for both males and females are presented in Table 3.

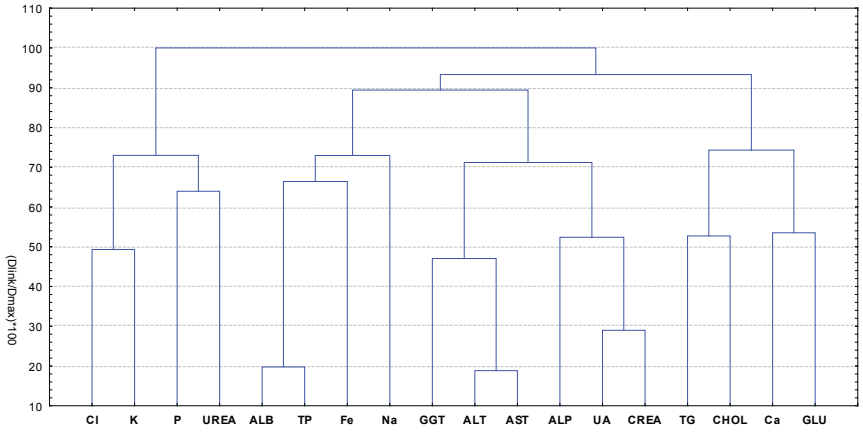
According to the Pearson test ($r>0.4$), the correlated couples of parameters for the 108 patients with ESRF were: UREA/CREA (0.425, $p<0.01$); UREA/K (0.481, $p<0.01$); UREA/P (0.499, $p<0.01$); AST/ALT (0.718, $p<0.01$); and TP/ALB (0.686, $p<0.01$).

These correlations were used to identify groups of highly correlated biochemical variables. It is evident that the simple correlation analysis did not indicate specific

Table 3. Basic statistics for the tested biochemical parameters for male and female patients with ESRF (mean value, minimum and maximum values within a certain variable, standard deviation of the mean).

Male (ESRF) patients					Female (ESRF) patients				
Variable	Mean	Min	Max	SD	Variable	Mean	Min	Max	SD
GLU	110	51	449	54	GLU	112	51	238	39
UREA	181	43	301	44	UREA	162	28	263	39
CREA	9.4	0.7	14.8	2.8	CREA	8.2	0.4	13.2	2.6
UA	5.6	4.0	8.6	1.0	UA	5.4	0.9	7.5	1.2
CHOL	145	79	216	31	CHOL	158	73	293	42
TG	153	42	382	67	TG	171	45	388	93
AST	14.3	3.0	53.0	7.9	AST	18.7	5.0	119.0	16.8
ALT	31.3	21.0	72.0	7.6	ALT	33.9	15.0	130.0	18.7
ALP	81	38	281	38	ALP	107	25	437	87
GGT	40.2	15.0	205.0	30.6	GGT	51.2	9.0	388.0	75.9
Na	136	131	141	2	Na	137	129	148	3
K	5.2	3.3	7.6	0.9	K	5.3	4.2	7.0	0.7
Cl	102	95	109	3	CL	104	94	113	3
Ca	9.3	8.1	10.6	0.6	Ca	9.1	6.5	12.0	1.0
Fe	84	9	285	52	Fe	77	24	197	45
P	5.3	1.5	9.6	1.6	P	4.6	2.5	6.7	1.1
TP	6.8	4.1	8.4	0.6	TP	6.3	1.8	7.8	1.1
ALB	3.4	1.9	4.1	0.3	ALB	3.2	1.5	4.1	0.5

Figure 1. Hierarchical dendrogram of biochemical variables for all healthy individuals.



links between the biochemical parameters being studied.

In Figure 1, the hierarchical dendrogram for the clustering of the biochemical parameters for all investigated healthy subjects (N = 166) is plotted (complete method of linkage, squared Euclidean distance as similarity measure, standardization of the input data). For clustering, a total of 18 clinical parameters were chosen (Table 1).

From the hierarchical dendrogram of Figure 1, it can be concluded that the parameters are principally separated into two large clusters, each of them divided into subclusters as follows:

Cluster 1 (four parameters included): Cl, K, P and UREA.

Cluster 2 (fourteen parameters included): ALB, TP, Fe, Na, GGT, ALT, AST, ALP, UA, CREA, TG, CHOL, Ca, and GLU.

Subcluster 1: ALB, TP, Fe and Na; subcluster 2: GGT, ALT, AST, ALP, UA and CREA; subcluster 3: TG, CHOL, Ca and GLU.

Cluster analysis was also performed for the dataset consisting of the biochemical parameters of only healthy males, only healthy females, all patients with ESRF, only males with ESRF, and only females with ESRF. The respective hierarchical dendrograms for only healthy

Figure 2. Hierarchical dendrogram of biochemical variables for male healthy individuals

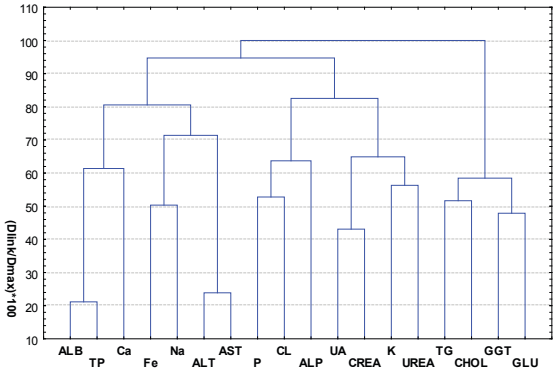


Figure 3. Hierarchical dendrogram of biochemical variables for female healthy individuals.

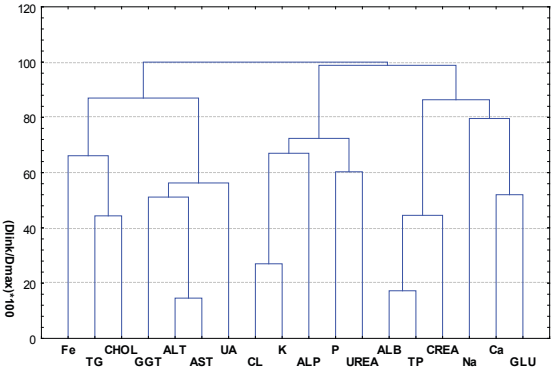


Figure 4. Hierarchical dendrogram of biochemical variables for all (ESRF) patients.

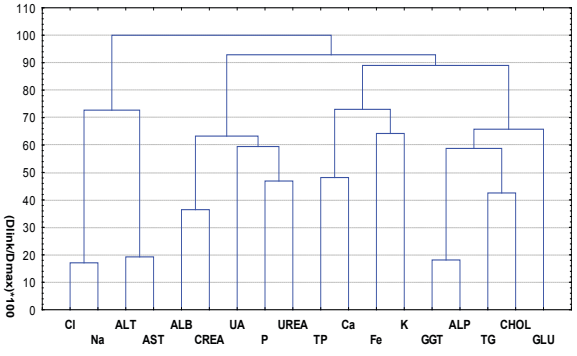


Figure 5. Hierarchical dendrogram of biochemical variables for male (ESRF) patients.

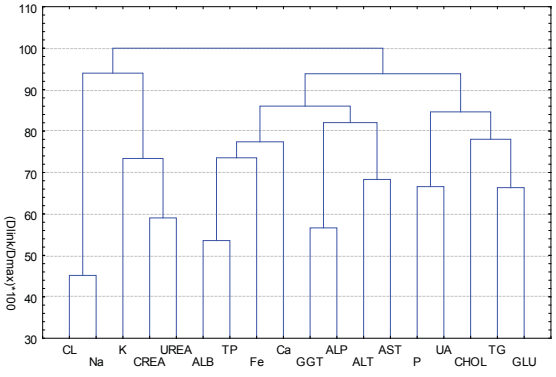
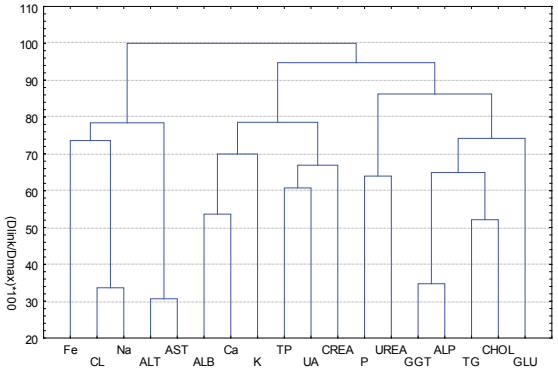


Figure 6. Hierarchical dendrogram of biochemical variables for female (ESRF).



males, only healthy females, all patients with ESRF, only males with ESRF, and only females with ESRF are shown in Figures 2-6.

Similarly, from the hierarchical dendrogram of Figure 4, it can be concluded that the parameters are principally separated into two large clusters, each of them divided into subclusters as follows:

Cluster 1 (four parameters included): Cl, Na, ALT and AST.

Cluster 2 (fourteen parameters included): ALB, CREA, UA, P, UREA, TP, Ca, Fe, K, GGT, ALP, TG, CHOL and GLU.

Subcluster 1: ALB, CREA, UA, P and UREA; subcluster 2: TP, Ca, Fe and K; and subcluster 3: GGT, ALP, TG, CHOL and GLU.

The usual classification approach of clustering is accompanied by factor analysis [principal components analysis (PCA) method], which is a typical projection and modelling approach. In general, the factor analysis confirms the results obtained by cluster analysis. The formation of seven latent factors, which are obviously responsible for the data structure, is proved for each of the different subsets: all healthy individuals, only healthy males, only healthy females, all patients with ESRF, only males with ESRF, and only females with ESRF.

The factor analysis (factor loading values) for all healthy individuals and for all patients with ESRF datasets is illustrated in Tables 5 and 7.

Table 4 provides the eigenvalues and the explanatory capability for the biochemical data of the set of all healthy individuals. The first and second factor, taken together, accounted for 33.90% of the total variability, whereas subsequent factors assisted in describing the biochemical data, but with a rapid diminishment in magnitude. The first seven factors each had an

Table 4. Individual and cumulative eigenvalues of all healthy subjects studied biochemical parameters.

No	Eigenvalue	Total Variance (%)	Cumulative (%)
1	3,340105	18,55614	18,55614
2	2,762778	15,34877	33,90491
3	1,697529	9,43072	43,33563
4	1,383295	7,68497	51,02060
5	1,325409	7,36338	58,38398
6	1,052105	5,84503	64,22901
7	1,047298	5,81832	70,04733

eigenvalue >1, and, together, accounted for 70.05% of the observed variation in biochemical data. Additional factors provided marginally less explanatory capability and were not examined further. Factor loadings reflect the correlations between the variables and the extracted factors. The factor loadings for the seven retained eigenvalues are shown in Table 5.

Table 6 provides the eigenvalues and the explanatory capability for the biochemical data of the set of all patients with ESRF. The first three factors, together, accounted for 43.58% of the total variability, whereas subsequent factors assisted in describing the biochemical data, but with a rapid diminishment in magnitude. The first seven factors each had an eigenvalue >1, and together, they accounted for 72.29% of the observed variation in biochemical data. The factor loadings for the seven retained eigenvalues are shown in Table 7.

Table 6. Individual and cumulative eigenvalues of all (ESRF) patients studied biochemical parameters.

No	Eigenvalue	Total Variance (%)	Cumulative (%)
1	3,148894	17,49385	17,49385
2	2,623711	14,57617	32,07003
3	2,071526	11,50848	43,57851
4	1,516469	8,42483	52,00334
5	1,292168	7,17871	59,18205
6	1,211290	6,72939	65,91144
7	1,147481	6,37489	72,28633

4. Discussion

As can be seen from Table 2, there were slight differences between the tested subgroups, e.g., higher averages for healthy males as compared to healthy females. There were also differences between male and female patients with ESRF (Table 3); for example GLU, CHOL, TG, AST, ALT, ALP, GGT, Na, K and Cl parameters had higher mean values in female than in male patients, and the rest of the variables had greater mean values in the male subgroup compared to those in the female subgroup.

The comparison between healthy individuals and patients with ESRF revealed significant differences between the two groups. More specifically, the patients had higher levels of serum UREA, CREA, TG, K, P, GGT, ALT and ALP; lower levels of serum AST, TP and ALB; and similar levels of serum Cl, Ca, Fe, UA, CHOL,

Table 5. Factor loadings of all healthy subjects studied biochemical parameters.

Variable	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6	Factor 7
GLU	-0,170764	0,070725	0,057579	-0,179320	0,860904	0,046765	0,013420
UREA	-0,058119	0,028801	0,102571	0,023976	0,064269	0,854204	-0,021350
CREA	0,589859	0,150513	0,273731	-0,315124	0,012662	0,423971	-0,069663
UA	0,430803	0,482191	0,190828	-0,343712	0,149396	0,335915	-0,062451
CHOL	-0,059223	0,069283	-0,428868	-0,609314	-0,031903	0,047536	0,262044
TG	0,034386	0,104007	-0,007138	-0,782219	0,137390	0,038074	0,009688
AST	0,067269	0,863484	-0,151596	0,086734	-0,079008	-0,043160	0,071987
ALT	-0,040114	0,897659	-0,023665	-0,145638	-0,015211	-0,011431	0,091939
ALP	0,322808	0,375172	0,184624	0,161038	0,248655	0,088152	-0,484674
GGT	-0,146218	0,696945	0,041375	-0,237412	0,327922	0,097622	-0,017308
Na	0,123928	-0,063414	-0,228337	0,490123	0,033725	0,367829	0,177911
K	0,263859	0,102066	0,751252	-0,013000	0,117036	0,173394	0,022527
CL	-0,063852	-0,167894	0,731656	0,038738	-0,171072	-0,001491	-0,040298
Ca	0,394539	0,013216	-0,335007	0,111335	0,606589	0,084180	-0,028797
Fe	0,310708	0,128405	0,033448	0,034484	-0,087541	0,179966	0,669618
P	0,217319	-0,164786	0,045530	0,044256	-0,382999	0,332481	-0,666122
TP	0,842051	-0,058521	-0,038537	0,007287	-0,093535	-0,123261	0,085527
ALB	0,903933	-0,041821	0,101359	0,129891	0,018963	0,031162	-0,008436

Table 7. Factor loadings of all (ESRF) patients studied biochemical parameters.

Variable	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6	Factor 7
GLU	0,118387	0,037700	0,185718	-0,143922	0,341679	0,430262	0,437355
UREA	0,812356	-0,010036	0,039942	-0,016086	-0,231806	-0,154428	-0,023452
CREA	0,567557	0,062867	-0,133886	0,468063	-0,328154	0,044384	-0,105229
UA	0,589327	-0,172507	0,148050	0,259566	0,414863	0,193725	-0,019320
CHOL	-0,071714	-0,207575	-0,075367	-0,045922	0,089570	0,761818	0,067508
TG	0,077993	-0,160204	0,084472	0,157046	-0,154719	0,811278	0,061530
AST	-0,037350	0,044658	0,939847	-0,031678	0,088975	0,017708	-0,037266
ALT	-0,009553	-0,268381	0,876538	0,093125	-0,094450	0,025576	-0,117875
ALP	-0,027425	-0,897300	0,094589	0,038713	0,042529	0,169604	0,036115
GGT	-0,050574	-0,880116	0,083346	-0,032022	0,009632	0,179305	0,164421
Na	-0,133994	0,140440	0,155496	-0,154308	0,187353	0,023237	-0,886021
K	0,146157	-0,082316	-0,070391	0,031611	-0,723894	-0,079664	0,148301
CL	0,033409	0,092662	0,061192	-0,175503	-0,052533	-0,142697	-0,879791
Ca	-0,244067	0,075208	0,082807	0,755411	0,031936	-0,075426	0,097749
Fe	0,056371	0,262818	0,237505	0,121797	-0,600906	0,298110	-0,167430
P	0,692515	0,130468	-0,083735	-0,014005	0,013797	0,072720	0,183097
TP	0,129188	-0,098660	0,062378	0,767587	0,030611	0,038498	0,170729
ALB	0,326288	0,013985	-0,089480	0,712239	-0,193353	0,132890	0,032554

Na and GLU compared to those of healthy individuals (Student's t-test, $p<0.05$).

The careful consideration of the content of the clusters (Figure 1) offers some interesting conclusions about the data classification when all healthy individuals were considered. The first cluster included electrolytes (Cl and K) and blood-specific health parameters including P and UREA. The second cluster involves proteins (ALB and TP), blood components (Fe and Na), enzymes (GGT, ALT, AST and ALP), parameters related to metabolic excretion processes (UA, CREA), and important general health parameters including levels of glucose, calcium, cholesterol and triglycerides.

This data analysis suggests how single biochemical parameters should be compared and related to each other if treatment is based on the combined values of the biochemical variables, not on individual ones. For instance, within a group of healthy persons, there is a stronger relation between the group of enzyme parameters (ALT, AST and GGT) and parameters like ALP, UA and CREA than to parameters like GLU, Ca, TG and CHOL. Therefore, specific patterns of the classified biochemical parameters can be seen (for the group of all healthy individuals):

- Enzyme-specific (including ALT, AST, GGT, ALP and two other parameters like UA and CREA)
- General health indicator (including glucose, cholesterol, triglyceride and calcium levels)
- Blood-specific indicator (including UREA, P, K and Cl)
- Protein-specific (including ALB, TP and two other parameters like Fe and Na)

In principle, it appears possible to design a set of substantial clinical parameters or a very specific health indicator consisting of parameters chosen from each pattern or from an average of several included in a cluster that are similar statistically.

The grouping of the clinical parameters for only healthy males (Figure 2) reveals differences compared to that of all healthy persons. For example, there is a difference in the linkage of the subcluster (ALT, AST and GGT) to the group (ALP, UA and CREA) of the second large cluster formed when all healthy individuals are considered. This may mean that the parameters for healthy males, when ALT and AST values are taken into consideration, should be related to the blood-specific indicator values (Fe and Na) rather than to GGT, ALP, UA and CREA indicators. Again, several patterns are formed, which correspond in principle to the idea of enzyme-specific (including ALT and AST, and two specific blood parameters like Fe and Na), general health indicator (TG, CHOL, GLU and GGT), protein-specific (ALB, TP and Ca), blood-specific indicator (Cl, P and one enzyme ALP), and major component excretion (UA, CREA, K and UREA) groups. This seems to be a classification rule for biochemical parameters that is specific to healthy males.

If the data for only healthy females are clustered (Figure 3), there is again a change in arrangement of the classification patterns. The careful inspection of Figure 3

indicates the presence of a blood-specific group (Cl, K, P, ALP and UREA), closely related to protein-specific (ALB, TP, CREA) and general health indicator (Na, Ca and GLU) groups. The enzyme-specific group (ALT, AST, GGT and UA) is relatively stable in this situation and includes one parameter from the major component excretion group (like UA**)** that is closely related to a blood-specific indicator (TG, CHOL and Fe) group. This seems to be a classification rule for biochemical parameters that is specific to healthy females.

It appears obvious that the possible introduction of a more general indicator of health state has to be sex-specific. This is a confirmation of the finding about differences between the average values of the single biochemical parameters between healthy males and females.

When the clinical parameters for all patients with ESRF (Figure 4) are grouped, again several patterns are formed, which correspond in principle to the idea of enzyme-specific (AST, ALT, with two other parameters like Cl and Na), general health indicator (TG, CHOL, GLU, GGT and ALP), blood-specific (TP, Ca, Fe and K), and major component excretion (UA, UREA, CREA, P and ALB) groups. This seems to be a classification rule for biochemical parameters that is specific to patients with ESRF.

If the data for only males or females with ESRF are clustered, there are again significant changes in the arrangement of the classification patterns among all patients with ESRF, males with ESRF, and females with ESRF, as can be seen in Figures 4, 5 and 6. Again, when the biochemical data of only male patients are clustered, several patterns are formed that correspond in principle to the idea of blood-specific indicator (Cl, Na), protein-specific (TP, ALB, Ca, Fe), major component excretion (UREA, CREA, K), general health indicator (GLU, TG, CHOL, UA and P), and enzyme-specific (ALT, AST, GGT and ALP) groups. This seems to be a classification rule for biochemical parameters that is specific to healthy males. If only female data are clustered (Figure 6), there are again changes in arrangement of the classification patterns. For example, the careful inspection of Figure 6 indicates the presence of a blood-specific group (P, UREA) that is closely related to the enzyme-specific group; the enzyme-specific group resembles the corresponding group with the same nomination, when all patients with ESRF are considered; the general health indicator group is exactly the same for the corresponding group with the same nomination, when all patients with ESRF are considered. This seems to be a classification rule for biochemical parameters that is specific to females with ESRF.

When the two groups (all healthy vs all patients with ESRF) are compared, major differences can be seen in the clustering of the biochemical variables in the study. The major excretion indicator group includes (CREA, UREA and UA) and two other parameters like (P and ALB), when all patients with ESRF are considered. It should be noted that the same pattern does not emerge when all healthy individuals are considered.

By comparison, when all patients with ESRF are considered, the enzyme-specific (ALT, AST, ALP and GGT) group separates into two different cluster. Moreover, there are differences in the linkage of the subcluster (TG, CHOL) to the group of enzymes of the second big cluster formed when all healthy individuals are considered. That may mean that the parameters for patients with ESRF should be related to enzyme indicator values (GGT, ALP) rather than to ALT and AST indicators, when TG and CHOL values are taken into consideration. Similarly, differences in clustering of biochemical parameters are also evident when male healthy individuals are compared with male patients with ESRF (Figures 2 and 5) and when female healthy individuals are compared with female patients with ESRF (Figures 3 and 6).

It seems obvious that the possible introduction of a more general health state indicator has to be specific to healthy persons. This is a confirmation of the finding about differences between the average values of the single biochemical parameters among healthy males, healthy females, male patients with ESRF, and female patients (Tables 2 and 3).

When all healthy individuals are considered, the application of factor analysis reveals seven latent factors that explain 70.05% of the total variance of the system, which is an indication for the adequacy of the PCA model. The marked loadings for all healthy individuals (Table 5) are statistically significant, according to Malinowski's test. The first factor contains TP and ALB, with a strong positive correlation between them. There is also a lesser positive correlation of CREA with the first factor. This factor could be conditionally named "protein-specific" factor and corresponds to the cluster with the same nomination. It explains 18.56% of the total variance. The next level of total variance (over 15%) can be explained by the second latent factor, which indicates a high positive correlation (factor loadings values) for AST, ALT and GGT and resembles the corresponding subcluster in a dendrogram of healthy individuals. Therefore, it could be conditionally named "enzyme-specific" factor. Additionally, there is slightly less of a correlation of UA with the parameters of the second factor. The third latent factor explains a substantial part of the total variance (9.43%) and reveals a strong

positive relation between the clinical parameters K and Cl, which allows its conditional designation as "blood-specific" factor. The latent factor four explains 7.68% of the total variance, and due to the high negative loadings of CHOL and TG, is logically designated as a "general health indicator" factor. There is also slightly less of a correlation of Na with the fourth factor. The fifth latent factor, with high factor loadings for GLU and Ca, could be conditionally named a "general health indicator" factor and corresponds to the cluster with the same nomination. It explains over 7% of the total variance. The sixth latent factor explains a substantial part of the total variance (5.84%) and involves the clinical parameter UREA, which allows its conditional designation as a "major component excretion" factor. Finally, the latent factor seven explains almost 6% of the total variance and, because of the high loadings of Fe and P, logically is designated a "blood-specific indicator" factor. There is also a slight correlation of ALP with factor seven.

The same method of analysis for all patients with ESRF (Table 7) reveals that the first latent factor with high positive factor loadings for UREA, CREA, UA and P could be conditionally named a "major component excretion" factor, corresponds to the cluster with the same nomination, and explains over 17% of the total variance. The next level of total variance (over 14%) can be explained by the second latent factor, which indicates a high negative correlation (factor loadings values) for ALP and GGT. Therefore, it could be again conditionally named an "enzyme specific" factor. The third latent factor explains a substantial part of the total variance (11.51%) and reveals a strong positive relation between the clinical parameters AST and ALT, which allows its conditional designation as "enzyme-specific" factor. The latent factor four explains over 8% of the total variance and, because of the high positive loadings of TP, ALB and Ca, logically is designated as "protein specific" factor. The fifth latent factor explains a substantial part of the total variance (over 7%) and reveals a strong positive relation between the clinical parameters K and Fe, which allows its conditional designation as a "blood-specific indicator" factor. Latent factor six explains 6.73% of the total variance and, because of the high positive loadings of CHOL and TG, logically is designated as a "general health indicator" factor. Moreover, there is a slight positive correlation of GLU with factor six. Finally, latent factor seven explains 6.37% of the total variance, and because of the high loadings of Na and Cl, logically is designated as a "blood specific indicator" factor. Additionally, there is again a slight correlation of GLU with factor seven.

It is readily apparent that the major groups of biochemical parameters interpreted by cluster analysis are also involved in the factor loadings presented in Tables 5 and 7. Thus, the classification scheme developed by cluster analysis is confirmed by factor analysis. This confirmation is an important indication that the biochemical laboratory parameters tested are, in fact, related and form groups of similar indicative properties.

In conclusion, the application of a typical classification approach such as cluster analysis to data sets consisting of the biochemical parameters of all healthy individuals and the same parameters of all patients with ESRF proved that there were points of similarity among all 18 biochemical parameters that formed major groups; these groups corresponded to the authors' assumption of the existence of several overall patterns of biochemical parameters that may be termed "enzyme-specific"; "general health indicator"; "major component excretion"; "blood-specific indicator"; and "protein-specific". These patterns also appeared when the general dataset was subdivided into "male" and "female" categories.

The performance of another technique of multivariate data analysis, factor analysis, proved the validity of a similar assumption. This projection and modelling method indicated the existence of seven latent factors, which explained 70.05% of the total variance of the system for all healthy individuals and more than 72% of the total variance of the system for all patients with ESRF, and, thus this analysis was responsible for the overall data structure.

All these results support the idea that a general health indicator could probably be constructed by taking into account the existing classification groups in the list of biochemical parameters.

Acknowledgements

This study is co-funded by the Research and Educational Committee, Technical and Educational Institute of Larissa. Greece.

References

[1] Martin M.V., Barroso S., Herraez O., De Sande F., Caravaca F., Cystatin C as a renal function estimator in advanced chronic renal failure stages, Nefrologia, 2006, 26, 433-438

[2] White C., Akbari A., Hussain N., Dinh L., Filler G., Lepage N., et al., Estimating Glomerular Filtration Rate in Kidney Transplantation: A Comparison between Serum Creatinine and Cystatin C-Based Methods, J. Am. Soc. Nephro.l, 2005, 16, 3763-3770

[3] Soveri I., Renal Dysfunction and Cardiovascular Disease, Acta Universitatis Upsaliensis, Uppsala, 2006

[4] Zappitelli M., Joseph L., Gupta I.R., Bell L., Paradis G., Validation of child serum creatinine-based prediction equations for glomerular filtration rate, Pediatr. Nephrol., 2007, 22, 272-281

[5] Tabak M.A., Christenson P.C., Fine R.N., Prediction of the Progression of Chronic Renal Failure in Children: Are Current Models Accurate?, Pediatrics, 1986, 6, 1007-1012

[6] Walser M., Drew H.H., Guldan J.L., Prediction of glomerular filtration rate from serum creatinine concentration in advanced chronic renal failure, Kidney Int., 1993, 44, 1145-1148

[7] Kuan Y., Hossain M., Surman J., El Nahas M., Haylor J., GRF prediction using the NDRD and Cockroft and Gault equations in patients with end-stage renal disease. Nephrol. Dial. Transplant., 2005, 20, 2394-2401

[8] Fontsere N., Bonal J., Navarro M., Riba J., Fraile M., Torres F., Romero R., A comparison of prediction equations for estimating glomerular filtration rate in adult patients with chronic kidney disease stages 4-5, Nephron Clin. Pract., 2006, 104, c160-c168

[9] Grubb A., Numan U., Bjork J., Lindstrom V., Rippe B., Sterner G., Christensson A., Simple cystatin C-based prediction equation for glomerular filtration rate compared with the modification of diet in renal disease prediction equation for adults and the Schwartz and the Counahan-Barratt prediction equations for children, Clin. Chem., 2005, 51, 1420-1431

[10] Johnston N., Low-Density Lipoprotein Oxidation and Renal Dysfunction, Acta Universitatis Upsaliensis, Uppsala, 2006

[11] Rossing P., Prediction, progression and prevention of diabetic nephropathy. The Minkowski Lecture 2005, Diabetologia, 2006, 49, 11-19

[12] Zoccali C., Mallamaci F., Tripepi G., Parlongo S., Cutrupi S., Benedetto F.A., et al. Norepinephrine and concentric hypertrophy in patients with end-stage renal disease, Hypertension, 2002, 40, 41-46

[13] Forestieri P., Formato A., Pilone V., Romano A., Monda A., Rhabdomyolysis after gastrectomy: increase in muscle enzymes does not predict fatel outcome, Obes. Surg., 2008, 18, 349-351

[14] Svensson M., Sundkvist G., Arnqvist H.J., Bjork E., Blohme G., Bolinder J., et. al., Signs of nephropathy may occur early in young adults with diabetes despite modern diabetes management, Diabetes Care, 2003, 26, 2903-2909

[15] Dyachenko P., Monselise A., Shustak A., Ziv M., Rozenman D., Nail disorders in patients with chronic renal failure and undergoing haemodialysis treatment: a case-control study, J. Eur. Acad. Dermatol. Venereol., 2007, 21, 340-344

[16] Srnak MJ., Levey AS., Schoolwerth A.C., Coresh J., Culleton B., et. al., Kidney disease as a risk factor for development of cardiovascular disease, Circulation, 2003, 108, 2154-2169

[17] Malyszko J., Malyszko JS., Pawlak K., Mysliwiec M., Hepcidin, iron status, and renal function in chronic renal failure, kidney transplantation, and hemodialysis, Am. J. Hematol., 2006, 81, 832-837

[18] Iseki K., Uehara H., Nishime K., Tokuyama K., Yoshihara K., et. al.. Impact of the initial levels of laboratory variables on survival in chronic dialysis patients. Am. J. Kidney Dis., 1996, 4, 541-548

[19] Geddes C.C., Van Dijk C.W., McArthur S., Metcalfe W., Jager K., et. al., The ERA-EDTA cohort-comparison of methods to predict survival on renal replacement therapy, Nephrol. Dial. Transplant., 2006, 21, 945-956

[20] Akgul A., Bilgic A., Ibis A., Ozdemir FN., Arat Z., Haberal M., Is uric acid a predictive factor for Graft dysfunction in renal transplant recipients?, Transplantation Proceedings, 2007, 39, 1023-1026

[21] Musso C., Liakopoulos V., De Miguel R., Imperiali N., Algranati L., Transtubular potassium concentration gradient: comparison between healthy old people and chronic renal failure patients, Int. Urol. Nephrol., 2006, 38, 387-390

[22] Gilbertson D.T., Liu J., Xue J.L., Louis T.A., Solid C.A., Ebben J.P., et al., Projecting the number of patients with end-stage renal disease in the United States to the year 2015, J. Am. Soc. Nephrol., 2005, 16, 3736-3741

[23] Perazella M.A., Khan S., Increased mortality in chronic kidney disease: a call to action, Am. J.

Med. Sci., 2006, 331, 150-153
[24] Sarnak M.J., Levey A.S., Schoolwerth A.C., Coresh J., Culleton B., et. al., Kidney disease as a risk factor for development of cardiovascular disease, Circulation, 2003, 108, 2154-69
[25] Tonelli M., Wiebe N., Culleton B., House A., Rabbat C., Fok M., et al., Chronic kidney disease and mortality risk: a systematic review, J, Am. Soc. Nephrol., 2006, 17, 2034-47
[26] Papaioannou A., Simeonov V., Plageras P., Dovriki E.,Spanos T., Multivariate statistical interpretation of laboratory clinical data, Central European J. Med., 2007, 3, 319-334
[27] Slockbower JM., Blumenfeld TA. (ed), Collection and handling of Laboratory Speciments, Philadelphic: Lippincott Co, 201, 1983
[28] NCCLS: “Procedures for the Collection of Diagnostic Blood Specimens by Venipuncture”, NCCLS, Document H3-A3 Wayne, PA: NCCLS, 1991
[29] NCCLS: “Procedures for the Collection of Diagnostic Blood Specimens by Skin Puncture”, NCCLS Document H4-A3 Wayne, PA: NCCLS, 1991
[30] NCCLS: “Internal Quality Control Testing: Principles and Definitions”, NCCLS, Document C24-A Wayne, PA: NCCLS, 1991
[31] Kafka M.T., Internal quality control, proficiency testing and the clinical relevance of laboratory testing, Arch. Pathol. Lab. Med., 1988, 112, 449-53
[32] Deming S., Chemometrics: an Overview, Clin. Chem., 1986, 32, 1702-1706
[33] Vogt W., Nagel D., Cluster Analysis in Diagnosis, Clin. Chem., 1992, 38, 182-198
[34] Vanderginste B., Massart DL., Buydens L., De Jong S., Lewi P., Smeyers-Verbeke J., Handbook of Chemometrics and Qualimetrics, Elsevier, Amsterdam, 1998
[35] Massart D.L,.Kaufman L., The Interpretation of analytical chemical data by the Use of Cluster Analysis, J. Wiley, New York, 1983
[36] Winkel P., Patterns and Clusters-Multivariate Approach for Interpreting Clinical Chemistry Results, Clin. Chem., 1973, 19, 1329-1338
[37] Grams R., Lezotte D., and Gudat J., Establishing a Multivariate Clinical Laboratory Data Base, J. Med. Sys.. 1978, 2, 355-362
[38] Nguyen H., Altinger J., Carrieri-Kohlman V., Gormley J., Paul S., Stulbarg M., Factor Analysis of Laboratory and Clinical Measurements of Dyspnea in Patients with Chronic Obstructive Pulmonary Disease, Journal of Pain and Symptom Management, 2003, 25, 118-127
[39] Plomteux G., Multivariate Analysis of an Enzymic Profile for the Differential Diagnosis of Viral Hepatitis, Clin. Chem., 1980, 26, 1897-1899
[40] Poupard J., Gagnon B., Stanhope M., Stewart C., Methods for Data Mining from Large multinational Surveillance Studies, Antimicrobial Agents and Chemotherapy, 2002, 46, 2409-2419
[41] Chang C.C., Cheng C.S., A structural design of clinical decision support system for chronic diseases risk management, Central European J. Med., 2007, 2, 129-139