INTRODUCTION

Aim: This study performs an Affinity Analysis on diagnosis and prescription data in order to discover co-occurrence relationships among diseases and pharmaceutical active ingredients prescribed to different patient groups.

Methods: Association rule and sequential pattern mining techniques have been applied. Rules and patterns are related both to the diagnosis and pharmaceutical active ingredients that have been prescribed to patients.

METHODOLOGY

A. Data collection

Patient sub-groups:
- HT: 1,763 patients (1,028 women, 734 men) who have been diagnosed with hypertension at least once in visit 10,136 distinct visit records.
- HCl: 1,295 patients (805 women, 490 men) who have been diagnosed with hypercholesterolemia at least once. 7,855 distinct visit records.
- HCt: 729 patients (448 women, 281 men) who have been diagnosed with hypertension and hypercholesterolemia at least once. 5,370 distinct visit records.

B. Data pre-processing

- A total of 17,758 distinct visits of 6,184 distinct patients.
- Visit (ICD-10 codes, prescribed substances)

C. Data transformation

Transaction database:
- Each visit is a transaction, containing all the diagnosis and drugs prescribed in this visit.
- Each patient is a transaction, containing all the distinct diagnosis for the patient and all the distinct drugs prescribed to the patient.

Knowledge-extracted (associations):
- Associations that frequently co-occur among the patients' visits.
- Rules of medicines that patients take at least once in the monitored period.
- When ordered of diagnosis/drugs is important to sequential associations.

D. Data mining

Application of association rules and sequential association rules implemented on RapidMiner Studio.

Use a minimum support and minimum confidence threshold, which limits the number of generated rules and increases their importance.

E. Interpretation of the results

- The distribution of cases per disease.
- The list of the most popular diseases.
- The distribution of cases per age group and gender for the most popular diagnoses and the most prescribed medications.

The World Health Organization's classification of age groups has been employed.

RESULTS

The rates have been projected to 300,000 persons, using the Greek Statistics Authority (GSA) data for the population of the city of Athens, which had 3.5 million inhabitants according to the 2011 census.

CONCLUSIONS

Results uncovered a correlation of hypertension and hypercholesterolemia with:
- coronary artery disease, gastroesophageal reflux disease and insulin dependent diabetes mellitus.

Additionally, in patients with hypertension we have observed positive correlation with:
- anxiety disorders.

Finally, in patients diagnosed with both hypertension and hypercholesterolemia, a correlation was observed with:
- non-insulin dependent diabetes,
- anxiety disorders,
- depressive disorders, and osteoporosis.