<u>Original Paper</u>

# Associations Among Multimorbid Conditions in Hospitalized Middle-aged and Older Adults in China: Statistical Analysis of Medical Records

Yan Zhang[1,2*], MPH; Chao Chen[1,3*], MSc; Lingfeng Huang[1,3*], MM; Gang Liu[4*], MM; Tingyu Lian[1,3], MPH; Mingjuan Yin[1,3], MM; Zhiguang Zhao[5], PhD; Jian Xu[2], PhD; Ruoling Chen[6], PhD; Yingbin Fu[4], BM; Dongmei Liang[1,3], MM; Jinmei Zeng[1,3], MPH; Jindong Ni[1,3], PhD

[1]Precision Key Laboratory of Public Health, School of Public Health, Guangdong Medical University, Dongguan, China

[2]Department of Elderly Health Management, Shenzhen Center for Chronic Disease Control, Shenzhen, China

[3]Institute of Public Health and Wellness, Guangdong Medical University, Dongguan, China

[4]Department of Primary Public Health Promotion, Shenzhen Center for Disease Control and Prevention, Shenzhen, China

[5]Administration Office, Shenzhen Center for Chronic Disease Control, Shenzhen, China

[6]Faculty of Education, Health and Wellbeing, University of Wolverhampton, Wolverhampton, United Kingdom

[*]these authors contributed equally

**Corresponding Author:**
Jindong Ni, PhD
Precision Key Laboratory of Public Health
School of Public Health
Guangdong Medical University
No.1 Xincheng Road, Songshan Lake
Dongguan, 523808
China
Phone: 86 15817668208
Email: nijd-gw@gdmu.edu.cn

## *Abstract*

**Background:** Multimorbidity has become a new challenge for medical systems and public health policy. Understanding the patterns of and associations among multimorbid conditions should be given priority. It may assist with the early detection of multimorbidity and thus improve quality of life in older adults.

**Objective:** This study aims to comprehensively analyze and compare associations among multimorbid conditions by age and sex in a large number of middle-aged and older Chinese adults.

**Methods:** Data from the home pages of inpatient medical records in the Shenzhen National Health Information Platform were evaluated. From January 1, 2017, to December 31, 2018, inpatients aged 50 years and older who had been diagnosed with at least one of 40 conditions were included in this study. Their demographic characteristics (age and sex) and inpatient diagnoses were extracted. Association rule mining, Chi-square tests, and decision tree analyses were combined to identify associations between multiple chronic conditions.

**Results:** In total, 306,264 hospitalized cases with available information on related chronic conditions were included in this study. The prevalence of multimorbidity in the overall population was 76.46%. The combined results of the 3 analyses showed that, in patients aged 50 years to 64 years, lipoprotein metabolism disorder tended to be comorbid with multiple chronic conditions. Gout and lipoprotein metabolism disorder had the strongest association. Among patients aged 65 years or older, there were strong associations between cerebrovascular disease, heart disease, lipoprotein metabolism disorder, and peripheral vascular disease. The strongest associations were observed between senile cataract and glaucoma in men and women. In particular, the association between osteoporosis and malignant tumor was only observed in middle-aged and older men, while the association between anemia and chronic kidney disease was only observed in older women.

**Conclusions:** Multimorbidity was prevalent among middle-aged and older Chinese individuals. The results of this comprehensive analysis of 4 age-sex subgroups suggested that associations between particular conditions within the sex and age groups occurred more frequently than expected by random chance. This provides evidence for further research on disease clusters and for health

care providers to develop different strategies based on age and sex to improve the early identification and treatment of multimorbidity.

**KEYWORDS**

multimorbidity; chronic conditions; aging; association rule mining; decision tree analysis

## *Introduction*

### Background

China is the world's most populous country and has the largest aging population. The population aged 65 years and older has markedly increased in recent years, and there were approximately 190 million people aged 65 years and older in China in 2020 [1]. With such a large aging population, chronic conditions are a major contributor to health burden, inequalities in health outcomes, and economic burden in China [2]. Multimorbidity (defined as 2 or more coexisting chronic conditions) has become a new challenge for medical systems and public health policy [3-5]. Multimorbidity is often associated with functional limitations, reduced quality of life, higher mortality, higher rates of adverse drug events, and frequent use of health services [6,7]. Despite the growing number of studies suggesting that multimorbidity is normal for older adults, the majority of health care systems and public health policies is focused on the treatment of individual diseases rather than a complex network of diseases [3]. The incidence of multimorbidity is latent, and the progression is slow [8]. If early detection and diagnosis are not efficient and timely, this not only will delay treatment and prognosis and affect the development of the disease but also may lead to premature death [9]. Therefore, understanding the patterns and associations among multimorbid conditions should be given priority, which may assist the early diagnosis of multimorbidity and thus improve quality of life of older adults [10].

An increasing number of studies have reported on the frequent combinations of diseases and described the patterns of multimorbidity. These studies used various methods, such as generating all possible combinations of chronic diseases, estimating observed-to-expected ratios or relative risk among the most common combination of 2 or 3 chronic conditions [11], cluster analysis [12,13], latent class analysis [14,15], factor analyses [16,17], and network analysis [4,18]. These methods are similar and investigate combinations of conditions but do not elucidate associations and the prioritization of associations between individual conditions. Furthermore, these disease combinations are mainly based on a single algorithm and lack further methods to verify their stability.

Association rule mining (ARM) is now being used to explore associations between frequent diseases [6]. ARM, a data mining technique used extensively in health care, attempts to identify and predict rules by extracting simple structures from a set of items in a database [19]. However, extrapolation of the association results based on existing samples and the priority of the associated condition of the target conditions are not taken into account in traditional ARM. With the addition of the Chi-square test and decision tree analysis, these disadvantages

can be avoided. The Chi-square test is a statistical method based on the difference in rate distribution, which can be used to test the statistical significance of the associations between the antecedent conditions and the consequent conditions in the association results, in order to extrapolate the sample results to the population situation. Decision tree analysis, a powerful statistical tool, has been successfully applied to recursively split independent variables into groups to predict an outcome [20,21]. In previous studies, it was also utilized to explore associated factors with survival in breast cancer patients [22], examine the interaction of shared variables to predict survival in patients with newly diagnosed malignant pleural mesothelioma [23], and investigate the prognostic importance of each factor for overall survival [24]. Unlike common methods, decision tree analysis can be used to classify factors to determine their importance to the target variables and decide which factor has the strongest association with the dependent variable at each point in the tree structure [25]. The combination of the 3 methods can obviously strengthen the evidence of the association between conditions, which enables accurate clinical decision support in practice. For more details on comparisons with currently used methods, please refer to Multimedia Appendix 1.

In addition, most studies on multimorbidity in China were conducted in community-dwelling populations, and self-reported questionnaires were used to define chronic diseases, which may have been affected by recall and reporting bias [11,26]. Hospital medical records describe the occurrence, development, diagnosis, and treatment of patients, and more objective clinical diagnoses are used to define multimorbidity. Obtaining the medical records of hospitalized patients to study multimorbidity could avoid recall or reporting bias. Furthermore, although multimorbidity is strongly associated with sociodemographic factors, few studies have focused on multimorbidity associations by age and sex.

### Objectives

To better understand the multimorbidity patterns in middle-aged and older people, this study used the novel method of combining ARM with a traditional statistical significance test and decision tree analysis to examine and compare associations among multimorbid conditions by age and sex in a large number of middle-aged and older Chinese adults using the home pages of inpatient medical records in Shenzhen, China. It was hoped that the results would provide possible potential trajectories between multimorbid conditions and improve population-specific approaches to early detection and management of multimorbidity.

XSL•FO

**RenderX**

## Methods

### Data Source

This study used data from the home pages of inpatient medical records in the Shenzhen National Health Information Platform, a data center that collects medical information on cases from all medical institutions in Shenzhen. The home pages of inpatient medical records, including information on hospitalized patients' demographic characteristics (age and sex), inpatient diagnoses, International Classification of Diseases version 10 (ICD-10) codes, and personal identifiers, were removed. All clinical visits by patients were linked to their unique encrypted identification number.

### Measurement of Multimorbidity and Study Population

In this study, the following 40 chronic conditions were selected based on the most frequently mentioned diseases in multimorbidity by previous studies that were considered to significantly impact long-term treatment and quality of life among middle-aged and older Chinese individuals [7,27]: hypertension (HT), diabetes mellitus (DM), lipoprotein metabolism disorder (LMD), chronic gastritis, chronic obstructive pulmonary disease, cerebrovascular disease (CBD), chronic kidney disease (CKD), spleen disease, peripheral vascular disease (PVD), varicose veins, schizophrenia, malignant tumor (MT), dementia, Alzheimer disease, bronchiectasis, glaucoma, senile cataract (SC), asthma, chronic nasopharyngitis, chronic viral hepatitis, thyroid disorders, hearing loss, dermatitis and eczema, anemia, migraine, chronic liver disease (CLD), depression, epilepsy, anxiety, Parkinson disease, sleep disorder, heart disease (HD), chronic gastric ulcer, gout, osteoporosis, transient cerebral ischemia, arthropathy, spondylosis, and dizziness/vertigo. Conditions were identified if they had been documented using inpatient ICD-10 codes in an individual's medical records. Multimedia Appendix 2 lists all chronic conditions included and their corresponding ICD-10 codes. For this study, multimorbidity was defined as having 2 or more concurrent chronic conditions.

In this study, patient inclusion criteria included (1) diagnoses with at least one of the aforementioned 40 conditions in all inpatient records from January 1, 2017, to December 31, 2018, and (2) aged 50 years or older on earlier records. Middle-aged patients with multimorbidity represent a large group, and the prevalence of multimorbidity ranges from 45% to 72% among middle-aged and older people older than 50 years [28]. Exclusion criteria were that none of the conditions were diagnosed in any inpatient records during the study period. A total of 306,264 patients were included.

### Statistical Analyses

#### Descriptive Statistics

First, descriptive statistics were used in the study population, including number, proportion (%), median, and IQR of age for sex (female and male). The top 10 prevalent chronic conditions with the largest composition ratio, including the average number (mean [SD]) of coexisting conditions, were evaluated. Furthermore, age was categorized into 2 subgroups (50-64 years and ≥65 years) and cross-combined with sex into 4 age-sex subgroups. The number and proportion were used to describe the distribution of patients with or without multimorbidity, and the Chi-square test was performed to compare differences in the characteristics of patients with and without multimorbidity.

### Association Rule Mining Based on Subgroups

To identify the associations between conditions by age and sex, 4 age-sex–based subgroup analyses were then performed. ARM was applied to determine common multimorbidity patterns that met a minimum requirement of measurement indicators. Association rules were relationships between sets of conditions from "antecedent" to "consequent" [29]. We used 3 common measurement indicators: (1) support (how frequently the condition combinations appear in the data set), (2) confidence (how frequently the consequent conditions occur, conditional on the antecedent conditions), and (3) lift (the ratio of the observed support to that expected if antecedent and consequent were independent) [30]. Lift was considered the main measure of significance in ARM. A lift of "1" means that the probabilities of occurrence of the antecedent and consequent are independent of each other. Hence, a higher lift indicates a higher chance of co-occurrence of the consequent with the antecedent and a more significant association [31]. Setting a higher threshold value would reduce the number of rules that might result in missing essential rules with low frequencies, and setting a lower threshold value could result in a large number of rules that might hinder the management from summarizing rules [29]. Thus, many rounds of testing and evaluation were carried out before defining final thresholds to mine reasonable rules and to ensure the robustness of the model performance. Considering the vast number of disease types in the data set, the rules satisfying support >1%, confidence >50%, and lift >1 were selected. All association rules were sorted by lifts, and the top 10 association rules with larger lifts in 4 subgroups were described.

### Chi-square Tests

To evaluate the statistical significance of the aforementioned association rules, Chi-square tests were applied. Odds ratios (ORs) and 95% CIs between antecedent conditions and consequent conditions in the association rules of the 4 age-sex subcategories are shown.

### Decision Tree Analysis

Furthermore, decision tree analysis was conducted to examine the conditions associated with the main consequent conditions in the association rules. Decision tree analysis examines the relationship between influencing factors and target variables [32]. The decision tree process is a nonparametric method that creates a tree-based classification model [33]. A decision tree contains 3 main parts: decision nodes, branches, and leaves. The internal variables of the model represent a tree structure in which a decision is made in each branch according to the data features [25]. The tree starts with a node and extends to the leaf. The risky paths are identified and shown in nodes [34]. In this study, we used decision tree analysis to determine the relationship between the conditions and main consequent conditions in rule results. Thus, the consequent conditions in rule results were used as target variables, while the remaining conditions were used as the independent variables. Splitting

criteria provides a rate for each predictor variable. Variables that have the best rate of splitting criteria are selected to remain in the model [25], which have a greater impact on the target variables, and in this study, various conditions were screened based on this feature. In the decision tree, the first variable or root node is the most important factor, and other variables can be classified in order of importance [35]. The decision trees were drawn to show the associated conditions with main consequent conditions in association rules.

The flowchart of the analyses is shown in Figure 1. All descriptive analyses and Chi-square tests were performed using SPSS version 25.0 (IBM Corp, Armonk, NY), with a .05 level of significance. ARM and decision tree analysis were carried out using R 3.4.0 (The R Foundation for Statistics and Mathematics, Vienna, Austria) with the arules package and the tree package. To make the results more intuitive, GraphPad Prism version 8.0 (GraphPad Software, San Diego, CA) was used to show the ORs and 95% CIs, and PowerPoint software 2021 version (Microsoft Corp, Redmond, WA) was used to draw decision trees.

**Figure 1.** The flowchart of the main research steps.
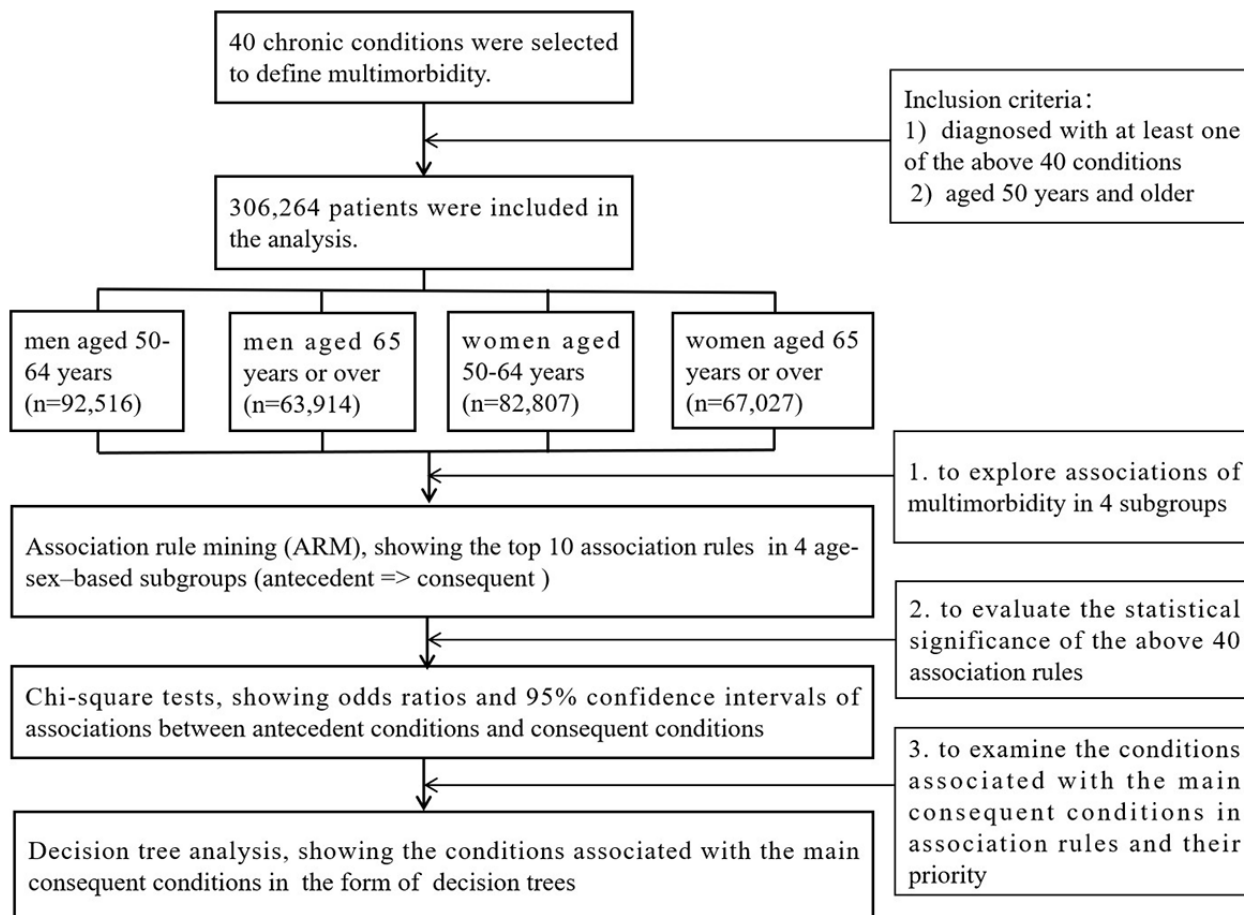


## Ethics Approval

This study was approved by the Institutional Review Ethics Committee of the Affiliated Hospital of Guangdong Medical University (YJYS202008). Informed consent was not required from participants as all data provided were deidentified.

## Results

### Characteristics of the Participants

In total, 306,264 hospitalized cases with available information on related chronic conditions were included in this study. The median age of the study population was 62 (IQR 55-71) years. There were more men than women, with men accounting for 51.08% (156,430/306,264) of the sample. The median age of the male and female participants was 62 (IQR 54-71) years and 63 (IQR 56-72) years, respectively (Table 1).

**Table 1.** Characteristics of the study population (N=306,264).

| Characteristics | All respondents | Men | Women |
|---|---|---|---|
| Number of people, n (%) | 306,264 (100) | 156,430 (51.08) | 149,834 (48.92) |
| Age (years), median (IQR) | 62 (55-71) | 62 (54-71) | 63 (56-72) |

## Characteristics of the Chronic Conditions

As shown in Table 2, 44.72% (136,972/306,264) of the study population had HT, which was the most prevalent condition. This was followed by HD (74,535/306,264, 24.34%), DM (70,917/306,264, 23.16%), CBD (68,151/306,264, 22.25%), LMD (65,385/306,264, 21.35%), CKD (63,470/306,264, 20.72%), CLD (61,829/306,264, 20.19%), PVD (51,311/306,264, 16.75%), spondylosis (42,982/306,264, 14.03%), and gout (33,984/306,264, 11.10%). Patients with these chronic conditions had an average multimorbidity burden of ≥4 chronic conditions per patient.

**Table 2.** Top 10 conditions with the largest composition ratio in all cases (N=306,264).

| Rank | Chronic conditions | Presence in all participants, n (%) | Number of co-occurring conditions, mean (SD) |
|---|---|---|---|
| 1 | Hypertension | 136,972 (44.72) | 4.79 (0.76) |
| 2 | Heart disease | 74,535 (24.34) | 4.97 (0.27) |
| 3 | Diabetes mellitus | 70,917 (23.16) | 4.84 (0.67) |
| 4 | Cerebrovascular disease | 68,151 (22.25) | 4.89 (0.53) |
| 5 | Lipoprotein metabolism disorder | 65,385 (21.35) | 4.91 (0.49) |
| 6 | Chronic kidney disease | 63,470 (20.72) | 4.81 (0.77) |
| 7 | Chronic liver disease | 61,829 (20.19) | 4.87 (0.59) |
| 8 | Peripheral vascular disease | 51,311 (16.75) | 4.95 (0.38) |
| 9 | Spondylosis | 42,982 (14.03) | 4.94 (0.39) |
| 10 | Gout | 33,984 (11.10) | 4.97 (0.29) |

## Differences in the Characteristics of Patients With and Without Multimorbidity

Of the 306,264 patients included, over 50% (175,323/306,264, 57.25%) were between 50 years and 64 years old (Table 3). The prevalence of multimorbidity in the overall population was 76.46% (234,156/306,264), with a higher prevalence in patients aged 65 years or older (108,937/306,264, 83.20%) than in those aged 50 years to 64 years (125,219/306,264, 71.42%). There were statistically significant sex differences in the prevalence of multimorbidity in the overall population, and patients aged 50 years to 64 years showed a higher prevalence in men than in women.

**Table 3.** Differences in the characteristics of patients with and without multimorbidity (N=306,264).

| Age groups | Multimorbidity, n (%) | No multimorbidity, n (%) | P value |
|---|---|---|---|
| **≥50 and ≤64 years** | | | |
| Men | 67,665 (73.14) | 24,851 (26.86) | <.001 |
| Women | 57,554 (69.50) | 25,253 (30.50) | |
| Total | 125,219 (71.42) | 50,104 (28.58) | —[a] |
| **≥65 years** | | | |
| Men | 53,305 (83.40) | 10,609 (16.60) | .052 |
| Women | 55,632 (83.00) | 11,395 (17.00) | |
| Total | 108,937 (83.20) | 22,004 (16.80) | — |
| **Overall sample** | | | |
| Men | 120,970 (77.33) | 35,460 (22.67) | <.001 |
| Women | 113,186 (75.54) | 36,648 (24.46) | |
| Total | 234,156 (76.46) | 72,108 (23.54) | — |

[a]Not applicable.

## Association Rules and Statistical Analysis Results

The top 10 association rules in 4 age-sex–based subgroups according to lifts are shown in Multimedia Appendix 3. Among men and women aged 50 years to 64 years, LMD tended to be comorbid with DM, CLD, gout, HT, and PVD, which occurred in 7 association rules in men and in 10 rules in women. In addition, the combination of osteoporosis and MT was observed to have the strongest association in men, with a lift of 6.60, whereas this combination was not found in women. For patients aged 65 years or older, PVD tended to be present in combination with HT, LMD, CBD, and HD, which occurred in 5 association rules in men and 4 in women among the top 10 rules, indicating that these antecedent combinations positively correlated with the occurrence of PVD. Furthermore, the strongest associations were observed between SC and glaucoma in men (lift=6.65) and in women (lift=4.93). In particular, the 4 association rules including osteoporosis and MT were only observed in men, and their lifts were all greater than 4, while the associations between anemia, gout, and CKD (lift=3.00) were only observed in women.

Statistical analysis (Chi-square tests) of the association rules in 4 age-sex–based subgroups was carried out, and the results are shown in Figure 2 and Multimedia Appendix 4. For all 40 rules, the ORs of the associations between antecedent conditions and consequent conditions were greater than "1," and the 95% CIs did not include "1," indicating that the latter conditions were more likely to be positive when the combinations of antecedent conditions were positive than negative.

**Figure 2.** Point estimates of the odds ratios and 95% CIs (1.96 SE) of the associations between antecedent conditions and consequent conditions in the association rules of 4 age-sex subgroups in (A) men aged 50-64 years, (B) women aged 50-64 years, (C) men aged 65 years or older, and (D) women aged 65 years or older. The * indicate significant findings. CBD: cerebrovascular disease; CG: chronic gastritis; CKD: chronic kidney disease; CLD: chronic liver disease; DM: diabetes mellitus; DV: dizziness/vertigo; HD: heart disease; HT: hypertension; LMD: lipoprotein metabolism disorder; MT: malignant tumor; PVD: peripheral vascular disease; SC: senile cataract; SD: spleen disease; TCI: transient cerebral ischemia.
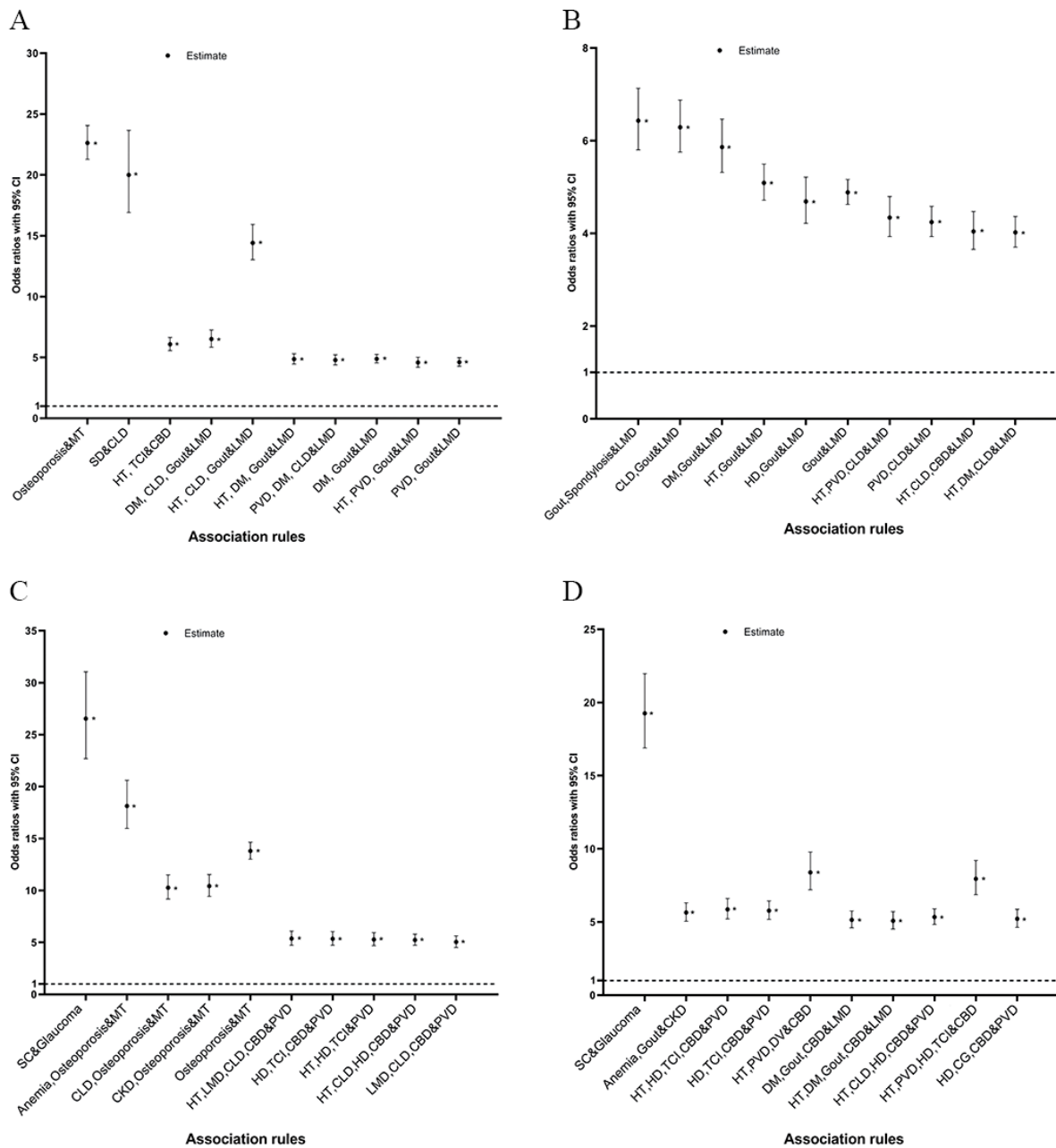


## Decision Tree Analysis of the Main Association Rules

Decision tree analysis was used to examine the associated comorbidities of the main consequent conditions in the rule results. The main decision trees are shown in Figure 3. Figure 3A shows that, in patients aged 50 years to 64 years, the decision tree with LMD as the dependent variable included nodes of gout, DM, HD, and CBD in men, and gout was at the top of the tree, indicating that 45.05% (5830/12,940) of patients with gout had LMD. More importantly, gout, CBD, and HD remained in the LMD decision tree for women. Gout was still at the top of

the tree, and 55.37% (3270/5920) of patients with gout had LMD. Figure 3B shows that, in men aged 50 years to 64 years and 65 years or older, osteoporosis remained at the top of the decision tree of MT in men, indicating that more than 50% of patients with osteoporosis had a comorbidity of MT. Among patients aged 65 years or older, condition nodes reserved in the decision tree of PVD included CBD, HD, and LMD in men and CBD, HD, LMD, and SC in women (Figure 3C). SC was the only node in the glaucoma decision tree in both sexes (Figure 3D). Furthermore, in women, SC, CLD, and anemia were observed in the decision tree of CKD (Figure 3E).

**Figure 3.** Decision trees with main consequent conditions as the target variables (other conditions divided into 2 subcategories: positive = "True"; negative = "False) in the association rules of different age-sex subgroups: (A) lipoprotein metabolism disorder (LMD) as the target variable in men and women aged 50-64 years, (B) malignant tumor (MT) as the target variable in men aged 50-64 years and 65 years or older, (C) peripheral vascular disease (PVD) as the target variable in men and women aged 65 years or older, (D) glaucoma as the target variable in men and women aged 65 years or older, (E) chronic kidney disease (CKD) as the target variable in women aged 65 years or older. All decimal values represent the proportion of the target conditions that were positive when the associated conditions were in the corresponding subgroup. CBD: cerebrovascular disease; CLD: chronic liver disease; DM: diabetes mellitus; HD: heart disease; HT: hypertension; SC: senile cataract.
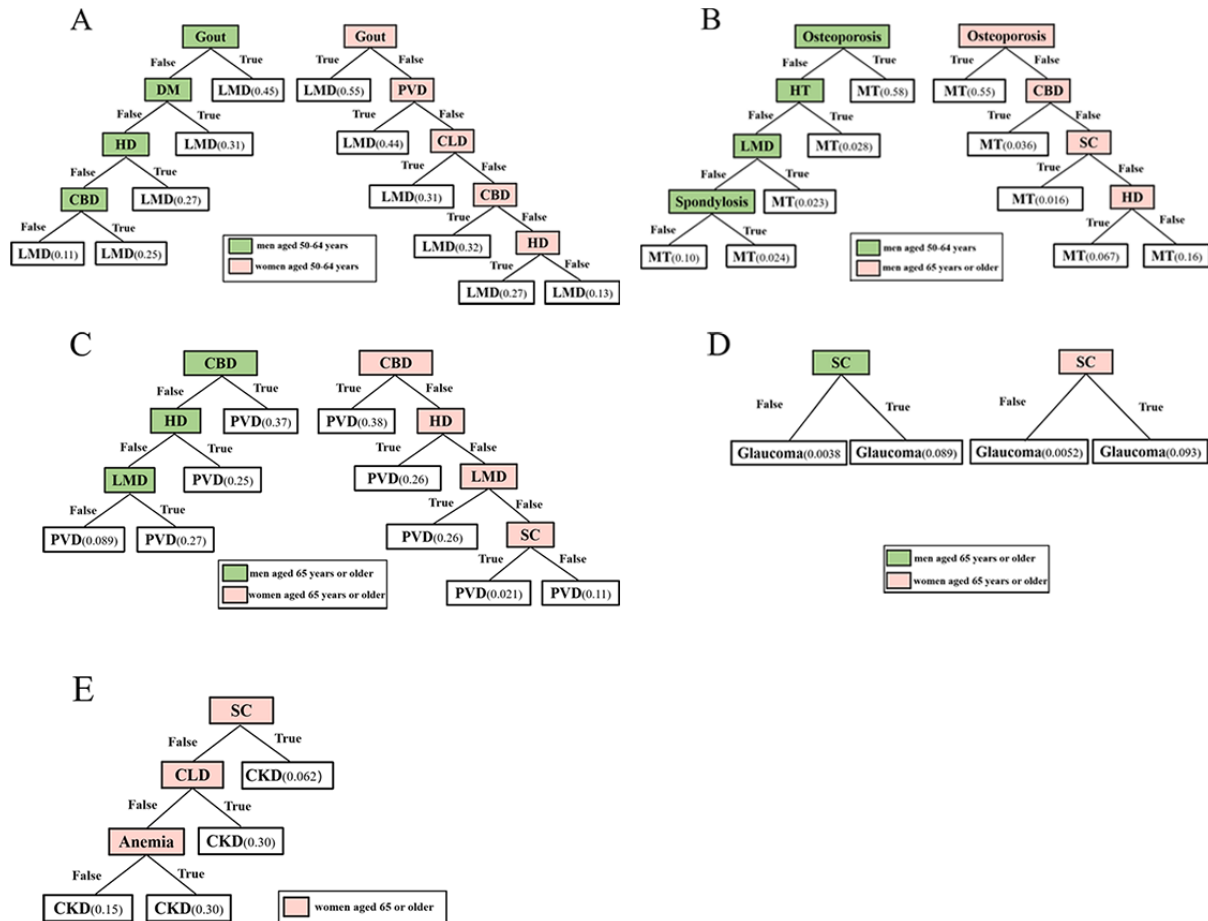


## Discussion

### Principal Findings

Understanding multimorbidity associations is an important public health priority for clinicians, academics, and funders alike [9]. This study was conducted to comprehensively evaluate the associations among multimorbid conditions based on the electronic hospitalized medical record home pages of a large sample of middle-aged and older Chinese people. To the best of our knowledge, this study is the first to evaluate multimorbidity associations using a comprehensive analysis with ARM, Chi-square tests, and decision tree analysis. Our analysis process not only revealed associations between particular conditions within different age-sex subgroups but also examined the importance of these associated chronic conditions for certain target conditions.

In this study of more 300,000 cases, more than 76% of patients were found to have 2 or more chronic conditions in the comprehensive list of 40 chronic conditions examined. The results showed that multiple conditions including HT, HD, DM, CBD, and LMD were the most common among hospitalized

middle-aged and older Chinese adults, and they co-occurred with more than 3 other conditions. This is similar to the findings reported in previous studies [4,7]. The prevalence of multimorbidity varied across the 2 age groups (50-64 years old and ≥65 years old) and both sex groups, reflecting the strong associations between multimorbidity and both age and sex [6]. Therefore, our subsequent analysis was based on specific age-sex subgroups to identify and compare the associations among multimorbid conditions within age and sex.

Association rules can reflect the interdependence and relevance between one condition and others. In our study, the ranked lift of most association rules indicated that LMD was the dominant condition among men and women aged 50 years to 64 years and was directly and indirectly associated with multiple conditions, including DM, CLD, gout, HT, and PVD or combinations of these conditions, which was also confirmed by statistical analysis. The potential mechanisms might include increased systemic inflammatory mediators and some adverse effects, such as physical inactivity, which are also risk factors for associated conditions [7]. Furthermore, in both men and women, gout appeared at the top of the decision tree with LMD as the dependent variable, which proved that the strong

association between gout and LMD was not coincidental. Our findings are consistent with those from previous studies. In a review, the author concluded that complex interconnections between gout and metabolic syndromes including LMD existed, showing that gout may play an important role in the manifestation of metabolic syndromes [36]. Therefore, proper management of one disease may have implications for early detection and prevention of another.

Among patients aged 65 years or older, ARM, statistical analysis, and decision tree analysis consistently found that PVD was closely interlinked with CBD, HD, and LMD. It was previously reported that these diseases share risk, pathophysiological, and prognostic features and their coexistence would cause a cumulative burden [27]. People with PVD are at significantly higher risk of myocardial infarction and stroke than the general population [37]. Although PVD can lead to adverse health outcomes, it has received little attention [38]. As an important comorbidity, PVD needs to be emphasized, and patients diagnosed with associated conditions should be targeted for PVD screening. Similarly, a significant association between SC and glaucoma was confirmed by all 3 methods, indicating that the probability of glaucoma was higher than the probability of other conditions when SC was present. This finding was consistent with that in a study based on large medical claims data among a Chinese population of 2 million [7]. The incidence of glaucoma and comorbid SC will increase with age, and measurements targeting those shared specific factors may benefit 2 or more related diseases [39].

Men aged 50 years to 64 years or 65 years or older reported a high prevalence of MT, with a high probability of co-occurrence with osteoporosis in the association rules and statistical analysis, which was consistent with the results of the decision tree analysis. Osteoporosis was found to be the condition most related with MT. Certain biological links have indeed been found between osteoporosis and MT, including the presence of important cytokines, hormones, and oxidative stress [40]. However, the sex difference between the 2 conditions in our research was inconsistent with some previous studies, which showed that osteoporosis and some types of MT, including breast cancer, thyroid cancer, and colorectal cancer, were more closely linked in women than in men [40-43]. This may be affected by factors such as MT type and age of the population, which requires further investigation in cancer subgroups. However, recognizing the existence of this association may help to guide the early screening of MT in Chinese middle-aged and older men with osteoporosis, especially the type with a high incidence in men.

The strong association between anemia, gout, and CKD was only detected in women aged 65 years or older by ARM and statistical analysis. The lift of 3.00 indicated that these conditions were 3 times as likely to occur simultaneously as they were alone. In the decision tree analysis, SC, CLD, and anemia were observed to be CKD-associated conditions. The common results of these 3 methods seemed to imply that there was a special association between anemia and CKD in this subgroup. Anemia is a common complication and contributes to increased morbidity and mortality in CKD patients, which

has been demonstrated previously [44,45]. A systematic review concluded that excess was a main contributor to the disordered iron homeostasis and anemia of CKD by impairing dietary iron absorption and iron mobilization from body stores [46]. Furthermore, possible explanations for this relationship only found in older women included shared risk factors of 2 conditions, such as aging and female sex [47,48]. Therefore, for older women, active improvement of anemia may be of great significance in preventing and delaying the development of CKD.

The main strength of this study is that a novel method was used, that is, the combination of ARM with a traditional statistical significance test and decision tree analysis, to examine the associations of multimorbidity. In particular, this was the first time that decision tree analysis was used in a multimorbidity study. Second, the disease diagnoses that defined multimorbidity in our analysis were based on a large sample of inpatient medical records, which avoided recall or reporting bias. Finally, our association analysis was based on age-sex subgroups, avoiding the confounding effects of age and sex. The present findings indicated that combinations of particular conditions within sex and age groups occur more frequently than expected by random chance. This provides evidence for further research on the potential mechanisms and risk factors for specific combinations and to encourage health care providers to develop population-specific approaches for early detection and management of multimorbidity according to sex and age.

## Limitations

Several limitations of our study must be acknowledged. First, our sample consisted of hospitalized cases, and mild and early cases may not have been included. In view of the fact that the research on multimorbidity in China is still at an early stage, our findings based on more severe cases may provide ideas for research on the early prevention of combinations of specific conditions. Second, we could not draw conclusions about causality effects between multiple conditions due to the cross-sectional design of the study. Finally, patients' socioeconomic status, family history, and lifestyle factors were not incorporated into the model in this analysis due to data availability, and the data set anonymized participants to avoid possible misuse; therefore, some potential confounding factors were not taken into consideration. However, given the advantages of our large sample size, the findings do provide support and a new perspective for future longitudinal or experimental studies to identify potential mechanisms and risk factors for specific combinations.

## Conclusions

Multimorbidity was prevalent among middle-aged and older Chinese individuals. The results of this comprehensive analysis of 4 age-sex subgroups suggested that associations among particular conditions within sex and age groups occurred more frequently than expected by random chance. This provides evidence for further research on disease clusters and for health care providers to develop different strategies, according to age and sex, to improve the early identification and treatment of multimorbidity.

## Acknowledgments

## Authors' Contributions

All authors contributed to this research. JN conceived and designed this research. YZ and CC extracted and analyzed the data and wrote the main part of the manuscript. LH and GL performed the main analysis and helped revise the manuscript. TL and MY searched the literature and designed the list of conditions. ZZ and JX provided great assistance with the whole process of research design and data analysis. RC and YF provided key ideas to write the paper. DL organized the tables, and JZ drew the figures. JN, YZ, CC, LH, and GL provided major suggestions for the revision of the paper. All authors edited and approved the final manuscript.

## Conflicts of Interest

None declared.

## Multimedia Appendix 1

Comparison of methods used in multimorbidity studies.
[DOCX File , 18 KB-Multimedia Appendix 1]

## Multimedia Appendix 2

ICD-10 numbers of 40 conditions included in the analysis.
[DOCX File , 15 KB-Multimedia Appendix 2]

## Multimedia Appendix 3

The top 10 association rules in the order of lifts.
[DOCX File , 33 KB-Multimedia Appendix 3]

## Multimedia Appendix 4

The results of the statistical analysis of association rules.
[DOCX File , 43 KB-Multimedia Appendix 4]

## References

1. China Statistical Yearbook 2021. National Bureau of Statistics of China. 2021. URL: http://www.stats.gov.cn/tjsj/ndsj/2021/indexeh.htm [accessed 2022-09-24]
2. Zhao Y, Atun R, Oldenburg B, McPake B, Tang S, Mercer SW, et al. Physical multimorbidity, health service use, and catastrophic health expenditure by socioeconomic groups in China: an analysis of population-based panel data. The Lancet Global Health 2020 Jun;8(6):e840-e849. [doi: 10.1016/s2214-109x(20)30127-3]
3. Park B, Ock M, Lee HA, Lee S, Han H, Jo M, et al. Multimorbidity and health-related quality of life in Koreans aged 50 or older using KNHANES 2013-2014. Health Qual Life Outcomes 2018 Sep 15;16(1):186 [FREE Full text] [doi: 10.1186/s12955-018-1016-6] [Medline: 30219061]
4. Hernández B, Reilly RB, Kenny RA. Investigation of multimorbidity and prevalent disease combinations in older Irish adults using network analysis and association rules. Sci Rep 2019 Oct 10;9(1):14567 [FREE Full text] [doi: 10.1038/s41598-019-51135-7] [Medline: 31601959]
5. Shi X, Lima SMDS, Mota CMDM, Lu Y, Stafford RS, Pereira CV. Prevalence of multimorbidity of chronic noncommunicable diseases in Brazil: population-based study. JMIR Public Health Surveill 2021 Nov 25;7(11):e29693 [FREE Full text] [doi: 10.2196/29693] [Medline: 34842558]
6. Lee Y, Kim H, Jeong H, Noh Y. Patterns of multimorbidity in adults: an association rules analysis using the Korea Health Panel. Int J Environ Res Public Health 2020 Apr 11;17(8):11278 [FREE Full text] [doi: 10.3390/ijerph17082618] [Medline: 32290367]
7. Wang X, Yao S, Wang M, Cao G, Chen Z, Huang Z, et al. Multimorbidity among two million adults in China. Int J Environ Res Public Health 2020 May 13;17(10):1 [FREE Full text] [doi: 10.3390/ijerph17103395] [Medline: 32414117]
8. Kim E, Lee T, Ji Y. Predictors of multimorbidity among Korean older adults: longitudinal secondary data analysis. Innovation in Aging 2021;5(S1):163. [doi: 10.1093/geroni/igab046.626]

XSL•FO

**RenderX**

9. Cezard G, McHale CT, Sullivan F, Bowles JKF, Keenan K. Studying trajectories of multimorbidity: a systematic scoping review of longitudinal approaches and evidence. BMJ Open 2021 Nov 22;11(11):e048485 [FREE Full text] [doi: 10.1136/bmjopen-2020-048485] [Medline: 34810182]

10. Sum G, Salisbury C, Koh G, Atun R, Oldenburg B, McPake B, et al. Implications of multimorbidity patterns on health care utilisation and quality of life in middle-income countries: cross-sectional analysis. J Glob Health 2019 Dec;9(2):020413 [FREE Full text] [doi: 10.7189/jogh.09.020413] [Medline: 31448114]

11. Zhang R, Lu Y, Shi L, Zhang S, Chang F. Prevalence and patterns of multimorbidity among the elderly in China: a cross-sectional study using national survey data. BMJ Open 2019 Aug 18;9(8):e024268 [FREE Full text] [doi: 10.1136/bmjopen-2018-024268] [Medline: 31427309]

12. Guisado-Clavero M, Roso-Llorach A, López-Jimenez T, Pons-Vigués M, Foguet-Boreu Q, Muñoz MA, et al. Multimorbidity patterns in the elderly: a prospective cohort study with cluster analysis. BMC Geriatr 2018 Jan 16;18(1):16 [FREE Full text] [doi: 10.1186/s12877-018-0705-7] [Medline: 29338690]

13. Marengoni A, Roso-Llorach A, Vetrano D, Fernández-Bertolín S, Guisado-Clavero M, Violán C, et al. Patterns of Multimorbidity in a Population-Based Cohort of Older People: Sociodemographic, Lifestyle, Clinical, and Functional Differences. J Gerontol A Biol Sci Med Sci 2020 Mar 09;75(4):798-805. [doi: 10.1093/gerona/glz137] [Medline: 31125398]

14. Olaya B, Moneta MV, Caballero FF, Tyrovolas S, Bayes I, Ayuso-Mateos JL, et al. Latent class analysis of multimorbidity patterns and associated outcomes in Spanish older adults: a prospective cohort study. BMC Geriatr 2017 Aug 18;17(1):186 [FREE Full text] [doi: 10.1186/s12877-017-0586-1] [Medline: 28821233]

15. Simões D, Araújo FA, Severo M, Monjardino T, Cruz I, Carmona L, et al. Patterns and Consequences of Multimorbidity in the General Population: There is No Chronic Disease Management Without Rheumatic Disease Management. Arthritis Care Res (Hoboken) 2017 Jan 17;69(1):12-20 [FREE Full text] [doi: 10.1002/acr.22996] [Medline: 27482954]

16. Menditto E, Gimeno Miguel A, Moreno Juste A, Poblador Plou B, Aza Pascual-Salcedo M, Orlando V, et al. Patterns of multimorbidity and polypharmacy in young and adult population: Systematic associations among chronic diseases and drugs using factor analysis. PLoS One 2019 Feb 6;14(2):e0210701. [doi: 10.1371/journal.pone.0210701] [Medline: 30726245]

17. Araujo MEA, Silva MT, Galvao TF, Nunes BP, Pereira MG. Prevalence and patterns of multimorbidity in Amazon Region of Brazil and associated determinants: a cross-sectional study. BMJ Open 2018 Nov 03;8(11):e023398 [FREE Full text] [doi: 10.1136/bmjopen-2018-023398] [Medline: 30391918]

18. Kalgotra P, Sharda R, Croff JM. Examining health disparities by gender: A multimorbidity network analysis of electronic medical record. Int J Med Inform 2017 Dec;108:22-28. [doi: 10.1016/j.ijmedinf.2017.09.014] [Medline: 29132627]

19. Shi X, Nikolic G, Van Pottelbergh G, van den Akker M, Vos R, De Moor B. Development of multimorbidity over time: an analysis of Belgium primary care data using Markov chains and weighted association rule mining. J Gerontol A Biol Sci Med Sci 2021 Jun 14;76(7):1234-1241 [FREE Full text] [doi: 10.1093/gerona/glaa278] [Medline: 33159204]

20. Guerrero MD, Vanderloo LM, Rhodes RE, Faulkner G, Moore SA, Tremblay MS. Canadian children's and youth's adherence to the 24-h movement guidelines during the COVID-19 pandemic: A decision tree analysis. J Sport Health Sci 2020 Jul;9(4):313-321 [FREE Full text] [doi: 10.1016/j.jshs.2020.06.005] [Medline: 32525098]

21. Aguiar FS, Almeida LL, Ruffino-Netto A, Kritski AL, Mello FC, Werneck GL. Classification and regression tree (CART) model to predict pulmonary tuberculosis in hospitalized patients. BMC Pulm Med 2012 Aug 07;12(1):40 [FREE Full text] [doi: 10.1186/1471-2466-12-40] [Medline: 22871182]

22. Ganggayah MD, Taib NA, Har YC, Lio P, Dhillon SK. Predicting factors for survival of breast cancer patients using machine learning techniques. BMC Med Inform Decis Mak 2019 Mar 22;19(1):48 [FREE Full text] [doi: 10.1186/s12911-019-0801-4] [Medline: 30902088]

23. Brims FJ, Meniawy TM, Duffus I, de Fonseka D, Segal A, Creaney J, et al. A novel clinical prediction model for prognosis in malignant pleural mesothelioma using decision tree analysis. J Thorac Oncol 2016 Apr;11(4):573-582 [FREE Full text] [doi: 10.1016/j.jtho.2015.12.108] [Medline: 26776867]

24. Shimose S, Kawaguchi T, Iwamoto H, Tanaka M, Miyazaki K, Ono M, et al. Controlling Nutritional Status (CONUT) score is associated with overall survival in patients with unresectable hepatocellular carcinoma treated with lenvatinib: a multicenter cohort study. Nutrients 2020 Apr 13;12(4):1076 [FREE Full text] [doi: 10.3390/nu12041076] [Medline: 32295043]

25. Vallée A, Safar ME, Blacher J. Application of a decision tree to establish factors associated with a nomogram of aortic stiffness. J Clin Hypertens (Greenwich) 2019 Oct 03;21(10):1484-1492 [FREE Full text] [doi: 10.1111/jch.13662] [Medline: 31479194]

26. Gu J, Chao J, Chen W, Xu H, Wu Z, Chen H, et al. Multimorbidity in the community-dwelling elderly in urban China. Arch Gerontol Geriatr 2017 Jan;68:62-67. [doi: 10.1016/j.archger.2016.09.001] [Medline: 27654809]

27. Wang HHX, Wang JJ, Wong SYS, Wong MCS, Li FJ, Wang PX, et al. Epidemiology of multimorbidity in China and implications for the healthcare system: cross-sectional survey among 162,464 community household residents in southern China. BMC Med 2014 Oct 23;12:188 [FREE Full text] [doi: 10.1186/s12916-014-0188-0] [Medline: 25338506]

XSL•FO

RenderX

28. Garin N, Koyanagi A, Chatterji S, Tyrovolas S, Olaya B, Leonardi M, et al. Global multimorbidity patterns: a cross-sectional, population-based, multi-country study. J Gerontol A Biol Sci Med Sci 2016 Feb 29;71(2):205-214 [FREE Full text] [doi: 10.1093/gerona/glv128] [Medline: 26419978]

29. Ma H, Ding J, Liu M, Liu Y. Connections between various disorders: combination pattern mining using apriori algorithm based on diagnosis Information from electronic medical records. Biomed Res Int 2022;2022:2199317 [FREE Full text] [doi: 10.1155/2022/2199317] [Medline: 35601156]

30. Ramezankhani A, Pournik O, Shahrabi J, Azizi F, Hadaegh F. An application of association rule mining to extract risk pattern for type 2 diabetes using tehran lipid and glucose study database. Int J Endocrinol Metab 2015 Apr 30;13(2):e25389 [FREE Full text] [doi: 10.5812/ijem.25389] [Medline: 25926855]

31. Held FP, Blyth F, Gnjidic D, Hirani V, Naganathan V, Waite LM, et al. Association rules analysis of comorbidity and multimorbidity: the Concord Health and Aging in Men Project. J Gerontol A Biol Sci Med Sci 2016 May 27;71(5):625-631. [doi: 10.1093/gerona/glv181] [Medline: 26508296]

32. Feng Y, Wang J, Shao Z, Chen Z, Yao T, Dong S, et al. Predicting related factors of immunological response to hepatitis B vaccine in hemodialysis patients based on integration of decision tree classification and logistic regression. Hum Vaccin Immunother 2021 Sep 02;17(9):3214-3220 [FREE Full text] [doi: 10.1080/21645515.2021.1895603] [Medline: 33989106]

33. Vallée A, Petruescu L, Kretz S, Safar M, Blacher J. Added value of aortic pulse wave velocity index in a predictive diagnosis decision tree of coronary heart disease. Am J Hypertens 2019 Mar 16;32(4):375-383. [doi: 10.1093/ajh/hpz004] [Medline: 30624553]

34. Amini P, Maroufizadeh S, Samani RO, Hamidi O, Sepidarkish M. Prevalence and determinants of preterm birth in Tehran, Iran: a comparison between logistic regression and decision tree methods. Osong Public Health Res Perspect 2017 Jun 30;8(3):195-200 [FREE Full text] [doi: 10.24171/j.phrp.2017.8.3.06] [Medline: 28781942]

35. Ahmed AM, Rizaner A, Ulusoy AH. A novel decision tree classification based on post-pruning with Bayes minimum risk. PLoS One 2018 Apr 4;13(4):e0194168 [FREE Full text] [doi: 10.1371/journal.pone.0194168] [Medline: 29617369]

36. Thottam GE, Krasnokutsky S, Pillinger MH. Gout and metabolic syndrome: a tangled web. Curr Rheumatol Rep 2017 Aug 26;19(10):60. [doi: 10.1007/s11926-017-0688-y] [Medline: 28844079]

37. Diehm C, Allenber J, Pittrow D. Mortality and vascular morbidity in older adults with asymptomatic versus symptomatic peripheral artery disease. Journal of Vascular Surgery 2010 Jul;52(1):253-254. [doi: 10.1016/j.jvs.2010.05.078]

38. Chen X, Zhou D, Shen J, Wu Y, Sun Q, Dong J, et al. Prevalence and causes of visual impairment in adults in Binhu District, Wuxi, China. Med Sci Monit 2018 Jan 16;24:317-323 [FREE Full text] [doi: 10.12659/msm.908218] [Medline: 29335399]

39. Zhao C, Cun Q, Tao Y, Yang W, Zhong H, Li F, et al. Effect of intraocular lens implantation on visual field in glaucoma and comorbid cataracts. Int J Ophthalmol 2020;13(4):580-586. [doi: 10.18240/ijo.2020.04.08] [Medline: 32399408]

40. Muhammad A, Mada SB, Malami I, Forcados GE, Erukainure OL, Sani H, et al. Postmenopausal osteoporosis and breast cancer: The biochemical links and beneficial effects of functional foods. Biomed Pharmacother 2018 Nov;107:571-582. [doi: 10.1016/j.biopha.2018.08.018] [Medline: 30114641]

41. Barzi A, Hershman DL, Till C, Barlow WE, Ramsey S, Lenz H, et al. Osteoporosis in colorectal cancer survivors: analysis of the linkage between SWOG trial enrollees and Medicare claims. Arch Osteoporos 2019 Jul 28;14(1):83 [FREE Full text] [doi: 10.1007/s11657-019-0629-7] [Medline: 31352608]

42. Papaleontiou M, Banerjee M, Reyes-Gastelum D, Hawley S, Haymart M. Risk of osteoporosis and fractures in patients with thyroid cancer: a case-control study in U.S. veterans. Oncologist 2019 Sep;24(9):1166-1173 [FREE Full text] [doi: 10.1634/theoncologist.2019-0234] [Medline: 31164453]

43. Shapiro CL. Osteoporosis: a long-term and late-effect of breast cancer treatments. Cancers (Basel) 2020 Oct 23;12(11):3094 [FREE Full text] [doi: 10.3390/cancers12113094] [Medline: 33114141]

44. Batchelor EK, Kapitsinou P, Pergola PE, Kovesdy CP, Jalal DI. Iron deficiency in chronic kidney disease: updates on pathophysiology, diagnosis, and treatment. JASN 2020 Feb 10;31(3):456-468. [doi: 10.1681/asn.2019020213]

45. Hanna RM, Streja E, Kalantar-Zadeh K. Burden of anemia in chronic kidney disease: beyond erythropoietin. Adv Ther 2021 Jan 29;38(1):52-75 [FREE Full text] [doi: 10.1007/s12325-020-01524-6] [Medline: 33123967]

46. Babitt JL, Lin HY. Mechanisms of anemia in CKD. JASN 2012 Aug 30;23(10):1631-1634. [doi: 10.1681/asn.2011111078]

47. Glassock RJ, Warnock DG, Delanaye P. The global burden of chronic kidney disease: estimates, variability and pitfalls. Nat Rev Nephrol 2017 Feb 12;13(2):104-114. [doi: 10.1038/nrneph.2016.163] [Medline: 27941934]

48. Stauder R, Valent P, Theurl I. Anemia at older age: etiologies, clinical implications, and management. Blood 2018 Feb 01;131(5):505-514 [FREE Full text] [doi: 10.1182/blood-2017-07-746446] [Medline: 29141943]

## Abbreviations

**ARM:** association rule mining
**CBD:** cerebrovascular disease
**CKD:** chronic kidney disease
**CLD:** chronic liver disease
**DM:** diabetes mellitus

XSL•FO
RenderX

**HD:** heart disease
**HT:** hypertension
**ICD-10:** International Classification of Diseases version 10
**LMD:** lipoprotein metabolism disorder
**MT:** malignant tumor
**OR:** odds ratio
**PVD:** peripheral vascular disease
**SC:** senile cataract