Social Medicine and Health Management (2025)
Clausius Scientific Press, Canada

DOI: 10.23977/socmhm.2025.060103
ISSN 2616-2210 Vol. 6 Num. 1

# *Analysis of Comorbidity Patterns of Chronic Diseases in Elderly Population in China Based on CHARLES Data*

**Meiyan Han**

*School of Clinical Medicine, Guangdong Medical University, Dongguan, 523808, Guangdong, China*
*1499600731@qq.com*

***Keywords:*** Comorbidity of chronic diseases, network analysis, negative binomial regression

***Abstract:*** This study aims to reveal the network pattern of comorbidity of chronic diseases and its main influencing factors, and provide scientific basis for comprehensive management of chronic diseases. The study used network analysis, hierarchical clustering and negative binomial regression model to analyze CHARLES large-scale questionnaire data, and systematically discussed the characteristics, patterns and mechanisms of health behaviors of chronic diseases. The results show that high-frequency co-occurring diseases (such as zdisease_1_and zdisease_14_) are concentrated in the center of the comorbid network, forming a strongly correlated core network; while diseases such as zdisease_12_appear as isolated nodes, indicating their unique etiology or low co-occurrence frequency. Hierarchical clustering further verified the close connection and local branch structure between diseases, and identified several small subgroups with high co-occurrence and bridging diseases (such as zdisease_10_). Negative binomial regression analysis revealed that health behaviors such as smoking, alcohol consumption, and physical activity significantly affected the number of comorbidities. For every unit increase in physical activity, the number of comorbidities decreases by approximately 2.5%; moderate drinkers have a significantly lower risk of comorbidities than non-drinkers, while the risk of comorbidities among current smokers is temporarily reduced due to compensation mechanisms. This study provides important reference for the prevention and control of chronic diseases.

## 1. Introduction

As the global aging process accelerates, the health problems of the elderly are receiving increasing attention. In China, the aging trend is particularly significant. According to data from the National Bureau of Statistics, the population aged 65 and above in China will exceed 200 million in 2021, accounting for 14.2% of the total population [1]. Among the elderly, the high incidence of chronic diseases and the comorbidity of multiple chronic diseases have become the main factors affecting their quality of life and medical burden.

Chronic diseases refer to diseases with long courses and slow progress, such as hypertension, diabetes, heart disease, etc. As we age, the probability of elderly people suffering from multiple chronic diseases increases significantly. This phenomenon is called chronic disease comorbidity. Studies have shown that comorbidity of chronic diseases not only increases the physical burden on

patients, but also increases the demand and cost of medical services and reduces the quality of life. In China, the current situation of chronic comorbidities among the elderly is not optimistic. A study based on 2018 data from China Health and Retirement Longitudinal Study (CHARLS) showed that among 10,836 elderly people aged 60 and above, 86.23% suffered from at least one chronic disease, and 65.14% suffered from two or more chronic diseases at the same time. Among them, hypertension (46.46%), arthritis or rheumatism (44.68%) and gastric or digestive system diseases (31.64%) are the three chronic diseases with the highest prevalence[1].

Understanding the patterns of chronic comorbidity among the elderly is crucial to formulating effective public health policies and medical service strategies. By analyzing comorbidity patterns, high-risk groups can be identified, medical resource allocation can be optimized, and targeted intervention measures can be formulated to improve the quality of life of the elderly and reduce medical burden.

However, at present, research on comorbidity patterns of chronic diseases in the elderly in my country is relatively limited. Some studies are limited to descriptive statistics and lack in-depth analysis of comorbidity patterns. In addition, there may be differences in the comorbidity patterns of chronic diseases among elderly people in different regions and different living habits, but these differences have not been fully studied.

The purpose of this study was to use CHARLS data to analyze the pattern of chronic comorbidity and its influencing factors among elderly people in China. Specific research questions include:

Is there a significant comorbid relationship between different chronic diseases?

What diseases form the core of the comorbid network?

Is there a bridging disease or an isolated disease?

What is the structure of the comorbid network among diseases?

Are there significant subgroups or clustering characteristics?

What are the effects of individual and environmental factors such as smoking, alcohol consumption, gender, marital status, medical insurance coverage, and regional differences on the number of comorbidities?

## 2. Literature review

The definition of "comorbidity" was first proposed by Professor Feinstein (1970) in the United States[2]. It refers to patients suffering from an index disease under study who also develop other diseases. Chronic comorbidity refers to the phenomenon that the same patient suffers from two or more chronic diseases at the same time. This phenomenon is particularly common among the elderly and has become an important factor affecting the health and quality of life of the elderly. According to the World Health Organization (WHO) definition, multi-disease refers to a person suffering from two or more diseases at the same time. Among the elderly, chronic comorbidities usually involve the coexistence of multiple chronic diseases such as hypertension, diabetes, heart disease, and arthritis [3]. Pan et al. (2023) found that common comorbidity patterns of chronic diseases among elderly people in China include the combination of hypertension and arthritis or rheumatism with a prevalence of 22.13%, and the combination of hypertension and gastric disease with a prevalence of 10.24%. In addition, the study also found that there are differences in comorbidity patterns of chronic diseases among the elderly population of different ages. Among young elderly people, the top three chronic disease comorbidity patterns with prevalence rates are hypertension and arthritis or rheumatism (19.81%), arthritis or rheumatism and gastric disease (10.14%), hypertension and gastric disease (9.94%).

With the aging of the global population, the phenomenon of comorbidity of chronic diseases (that is, the same patient suffers from two or more chronic diseases at the same time) has become

increasingly common and has become an important research topic in the field of public health. Scholars at home and abroad have conducted extensive research on the epidemic trends, pattern recognition, influencing factors and management strategies of chronic comorbidities.

Studies have shown that the prevalence of chronic comorbidities increases with age. In the Chinese mainland, the comorbidity rate of chronic diseases among the elderly is on the rise. A meta-analysis showed that the prevalence of chronic comorbidities among the elderly population in China was approximately 43.6% [4]. In addition, the comorbidity status varies among different regions and people with different socioeconomic status.

Research on comorbidity patterns of chronic diseases aims to identify common combinations of diseases to guide clinical practice and public health policy. Domestic studies use multiple methods to analyze comorbidity patterns. For example, some scholars have found that hypertension, diabetes and cardiovascular disease are common comorbidity combinations among the elderly through network analysis methods[5]. In addition, research based on disease synergy theory also provides a new perspective for identifying comorbid patterns[6].

Factors such as socioeconomic status, lifestyle, and mental health are considered to be closely related to comorbidities of chronic diseases. Studies have found that people with lower socioeconomic status are more likely to suffer from multiple chronic diseases [7]. In addition, unhealthy lifestyles, such as smoking, alcohol abuse, lack of exercise, etc., can also increase the risk of comorbidity[8].

Management strategies for patients with chronic comorbidities are one of the hot topics in current research. There are relatively mature comorbidity management models abroad, such as patient-centered comprehensive nursing models. Domestic research is also exploring management strategies suitable for China”s national conditions. For example, some scholars have proposed a community-based comprehensive management model to improve the quality of life of elderly patients with chronic comorbidity[9].

Socio-economic status usually includes indicators such as income, education level and occupation. Research by Wu et al. (2024) showed that there are differences in the prevalence of chronic diseases and comorbidities among middle-aged and elderly people with different social and economic status[10]. Middle-aged and elderly people with low per capita annual income, low education levels, and unemployed people with comorbidities. The higher risk of developing diseases may be related to factors such as the availability of medical resources, lack of health knowledge, and differences in living environments.

Studies have pointed out that chronic comorbidities and cognitive dysfunction share many common lifestyle risk factors (such as sedentary lifestyle, smoking, diet, etc.). Adopting an active and healthy lifestyle, such as increasing physical activity and improving diet structure, can effectively reduce the risk of chronic comorbidities[11].

Social support plays an important role in alleviating the negative impact of chronic comorbidities. Studies have shown that good social support can improve life satisfaction and reduce medical expenditures for elderly patients with comorbidity[12]. Therefore, social support factors should be considered when formulating management strategies for comorbidities of chronic diseases.

To sum up, comorbidities of chronic diseases have become an important challenge in the global public health field. Domestic and foreign research has made important progress in epidemic trends, pattern recognition, influencing factors and management strategies. However, further in-depth research is still needed, especially on the definition and classification of comorbidities in the elderly population, as well as the mechanism of the influence of socioeconomic status and lifestyle on comorbidities.

## 3. Research Methods

### 3.1. Data Sources

This study used data from China Health and Retirement Longitudinal Study (CHARLS). CHARLS is a national follow-up survey organized and implemented by the National Development Research Institute of Peking University. It mainly targets middle-aged and elderly people in China aged 45 and above and their spouses. CHARLS uses a multi-stage, stratified probability sampling method covering 28 provinces across the country. The data includes detailed information such as demographic characteristics, health status, lifestyle, and social and economic status. This study selected the latest public data in 2020, combined with the diagnosis and reporting of chronic diseases in the health module, and conducted analysis. There were a total of 19367 samples.

### 3.2. Study Design

Through the descriptive statistics method, the basic characteristics of various variables in the sample are first summarized, including factors or covariates such as gender, marriage, living area, medical insurance status, smoking, drinking, mental health status, exercise status, and the distribution of 15 common chronic diseases such as hypertension and heart disease. Statistical indicators include frequency, percentage, etc.

SPSS 17.0 software was then used to explore comorbid patterns of chronic diseases in the elderly using network analysis methods to ensure the accuracy and visualization of the results. Transform each interviewee”s chronic disease information into a disease matrix to express the comorbidity relationships between different diseases. An undirected weighted network is established based on the Co-occurrence Frequency, where nodes represent diseases and the weights of edges represent the comorbidity frequency of two diseases. The analysis content includes Network Density (indicating the strength of the overall comorbidity relationship), Degree (indicating the correlation of each disease in the comorbidity network), and Modularity Analysis (identifying comorbidity patterns. Clustering structure). In order to explore the impact of social and economic status, lifestyle and mental health on comorbidities of chronic diseases, this study used a generalized linear regression model to analyze the linear impact of the independent variable with the number of comorbidities as the dependent variable.

The CHARLS questionnaire covers 15 common chronic diseases, including hypertension, dyslipidemia (hyperlipidemia or hypoglycemia), diabetes or increased blood sugar (Including impaired glucose tolerance and elevated fasting blood sugar), cancer and other malignant tumors (excluding mild skin cancer), chronic lung disease (Such as chronic bronchitis, emphysema, pulmonary heart disease, etc., excluding tumors or cancer), liver diseases (Except for fatty liver, tumors or cancer), heart disease (such as myocardial infarction, coronary heart disease, angina pectoris, congestive heart failure and other heart diseases), stroke (stroke), kidney disease (excluding tumors or cancer), gastric disease or digestive system disease (excluding tumors or cancer), emotional and mental problems, memory-related diseases (such as Alzheimer”s disease, brain atrophy), Parkinson”s disease, arthritis or rheumatism, asthma (non-pulmonary disease). The definition of disease is based on respondents “self-reported diagnoses, combined with drug use information and medical records from questionnaires if necessary.

Independent variables include lifestyle and mental health. Lifestyle includes smoking (whether you currently smoke), drinking (whether you drink frequently in the past year), physical activity (classified as "high", "medium", and "low" based on the frequency of daily physical activity in the CHARLS questionnaire), and mental health includes depression levels.

Control variables include gender (male/female), region (divided into urban and rural areas), marital

status (married/divorced or widowed), and medical insurance (whether they participate in basic medical insurance).

Based on CHARLS data, this study comprehensively analyzed the current situation of chronic comorbidities among elderly people in China and its influencing factors through a research design combining descriptive statistics, network analysis and generalized linear regression model. The design of the research method aims to reveal the complex structure of comorbid relationships and provide scientific basis for the intervention and management of comorbid chronic diseases.

### 3.3. Model Construction

Based on the above design of independent variables and dependent variables, a generalized linear regression model for analyzing the number of comorbidities was constructed. Analysis of data characteristics found that the number of comorbidities ranged from 0 to 12, with a mean value of 1.84 and a standard deviation of 1.739. See Table 1.

Table 1: Statistical characteristics of the number of comorbidities.

| Number of comorbidities | | Statistics | | | |
|---|---|---|---|---|---|
| Effective | Missing | Mean | Variance Difference | Minima | Maxima |
| 19367 | 0 | 1.84 | 3.022 | 0 | 12 |

The data is right-biased, indicating that most of the values are concentrated on the low number of comorbidities (0-2), and few values are distributed on the high number of comorbidities (>6). The distribution is asymmetric, showing a typical distribution of counting data. See Figure 1.

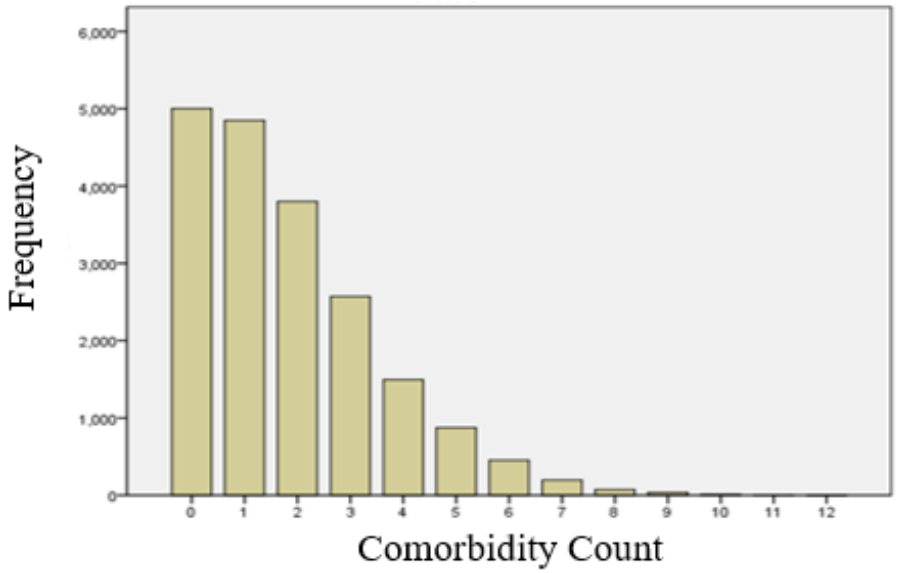


Figure 1: Bar chart of number of comorbidities.

Because the variance (3.023) is significantly greater than the mean (1.84), it means that the problem of excessive dispersion is significant. Based on this characteristic of data distribution, a negative binomial regression model is adopted because this model is suitable for cases where the dependent variable is count-type data and has excessive dispersion, and can handle the right-biased distribution, which is suitable for the analysis of the number of comorbidities.

For negative binomial regression, the dependent variable Y is a counting variable, and the model equation is:

$$\ln(E(Y))=\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k \quad (1)$$

Where: E(Y): is the expected value (i.e., the average value) of the dependent variable Y(i.e., Comfort Count). The "Comorbidity Count" here refers to the total number of different diseases that an individual suffers from at the same time. For example, if someone has high blood pressure and diabetes, their Comorbidity Count is 2. If someone has hypertension, diabetes, and coronary heart disease, the Comorbidity Count is 3. ln(E(Y)): is the logarithmic linking function of the dependent variable. β0: constant term (intercept). β1, β2, …, βk: Regression coefficient of independent variables. X1, X2, …, Xk: Predictors (factors or covariates), including Individual Income, Education, Gender, Region, Vital Status, Health Insurance, Smoking, Alcohol, Physical Activity, Depression.

After taking the logarithmic linking function to the exponent, the predicted value of the dependent variable can be obtained:

$$E(Y)=\exp(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k) \quad (2)$$

In terms of data selection, select da047 for "Smoking", select da051 for "Alcohol", select dc025 for "Depression", and select weighted days "Comprehensive da033_0_" for "Physical Activity": Weighted days = (days of high intensity da033_1_ * weight of high intensity 0.501) + (days of moderate intensity da033_1_ * weight of medium intensity 0.334) + (days of mild intensity da033_1_ * weight of light intensity 0.167). Gender selects ba001, Region selects ba008, Marital Status selects ba011, and Health Insurance selects ba017. Because there were too many missing values in the Individual Income (ga002) and Education (ba010) questionnaires, which seriously affected the fit of the model, it was not included in this analysis.

### 3.4. Coding Instructions

When entering the data into SPSS software for analysis in this study, most variables were coded using the original coding of the CHARLS Questionnaire Dictionary. A few variables needed to be merged from one data table into another data table for easy analysis, so they were re-coded to make it easy to identify during analysis. See Table 2 below for the codes used in all analyses.

Table 2: Coding table

| Variable Name | Encoding | Variable Name | Encoding |
|---|---|---|---|
| Hypertension | zdisease_1_ | Parkinson"s disease | zdisease_13_ |
| Dyslipidemia (hyperlipidemia or hypoglycemia | zdisease_2_ | Arthritis or rheumatism | zdisease_14_ |
| Diabetes or increased blood sugar (including impaired glucose tolerance and increased fasting blood sugar | zdisease_3_ | Asthma (non-pulmonary disease) | zdisease_15_ |
| Malignant tumors such as cancer (excluding mild skin cancer | zdisease_4_ | Gender | ba001 |
| Chronic lung diseases (such as chronic bronchitis, emphysema, pulmonary heart disease, etc., excluding tumors or cancer | zdisease_5_ | Marital Status | ba011 |
| Liver disease (excluding fatty liver, tumor or cancer | zdisease_6_ | Health Insurance | ba017 |
| Heart disease (such as myocardial infarction, coronary heart disease, angina pectoris, congestive heart failure and other heart diseases | zdisease_7_ | Region | ba008 |
| Stroke (Stroke) | zdisease_8_ | Smoking | da047 |
| Kidney disease (excluding tumors or cancers | zdisease_9_ | Alcohol | da051 |
| Stomach disease or digestive system disease (excluding tumors or cancer) | zdisease_10_ | Physical Activity | CCda033_0_ |
| Emotional and spiritual problems | zdisease_11_ | Depression | dc025 |
| Memory-related diseases (e.g. Alzheimer"s disease, brain atrophy) | zdisease_12_ | | |

## 4. Research Results

### 4.1. Descriptive Statistical Analysis Results

Through descriptive statistical analysis, the following results can be obtained: the sample is mainly middle-aged and elderly rural population, and chronic diseases are common, especially hypertension, arthritis and digestive system diseases. Lifestyle variables show that the proportion of smokers and drinkers is relatively low, at about one-third of each. See Table 3 below.

Table 3: Information on occurrence of various variables.

| Categorical variables | | Freq. | % | Types of CDs | Freq. | % |
|---|---|---|---|---|---|---|
| Gender | Male | 9090 | 46.87% | zdisease_1_ | 6563 | 33.9% |
| | Female | 10,305 | 53.13% | zdisease_2_ | 3726 | 19.2% |
| Marital Status | Married | 16352 | 84.3% | zdisease_3_ | 2165 | 11.2% |
| | Divorced/Widowed | 3043 | 15.7% | zdisease_4_ | 337 | 1.7% |
| Region | Urban | 7115 | 36.7% | zdisease_5_ | 2252 | 11.6% |
| | Rural | 12,270 | 63.30% | zdisease_6_ | 1137 | 5.9% |
| Health Insurance | Yes | 18461 | 95.2% | zdisease_7_ | 3268 | 16.9% |
| | No | 934 | 4.8% | zdisease_8_ | 1189 | 6.1% |
| Lifestyle Variables | | Freq. | % | zdisease_9_ | 1588 | 8.2% |
| | Smoker | 5738 | 29.6% | zdisease_10_ | 5303 | 27.4% |
| | Non-smoker | 13632 | 70.4% | zdisease_11_ | 384 | 2.0% |
| Alcohol | Drink Alcohol | 6824 | 35.2% | zdisease_12_ | 17 | 0.1% |
| | Non-drinker | 12546 | 64.8% | zdisease_13_ | 344 | 1.8% |
| | | | | zdisease_14_ | 6530 | 33.7% |
| | | | | zdisease_15_ | 887 | 4.6% |

### 4.2. Analysis Results of Comorbidity Patterns of Chronic Diseases

The network map of comorbid relationships is shown in Figure 2. The color gradient and the thickness of the line are used to represent the weight of the edge (co-occurrence frequency). The brighter the color and the thicker the line, the higher the co-occurrence frequency; the size of the node is used to represent the frequency of occurrence of the disease. The larger the node, the more times the disease appears in the sample.

You can see from the network distribution diagram: From the overall structure, most disease nodes are clustered in the central area of the network, indicating that there is a high frequency of co-occurrence among these diseases, and there may be certain associations or common risk factors. Zdisease_12_has fewer and finer connections with other diseases, indicating that its co-occurrence frequency is low and may be a highly isolated disease.

Judging from the characteristics of nodes, the sizes of nodes 1 (“zdisease_1_”) and 14 (“zdisease_14_”) are significantly higher than those of other nodes, indicating that these diseases have the highest frequency in the sample. They may be the dominant chronic diseases in the sample and deserve further attention. Although node 10 is slightly smaller in size than node 1 and node 14, it has a large number of connections, indicating that it may have the characteristics of bridging other diseases.

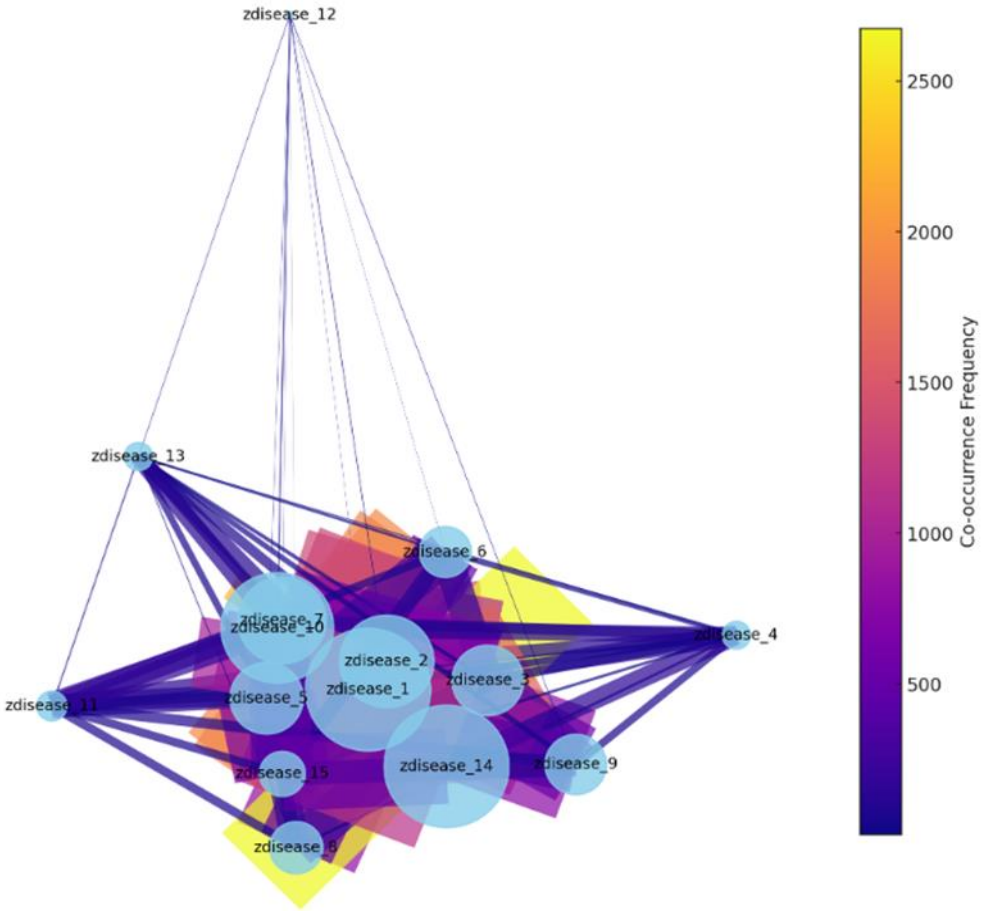


Figure 2: Co-Morbidity network.

Judging from the edge characteristics, the line between node 1 and node 14 is thick and bright, indicating that the co-occurrence frequency of these two diseases is very high. This strong association may indicate that they share common pathogenesis or risk factors. The frequency of co-occurrence between node 1 and node 10 is high. Node 14 also has a strong co-occurrence with node 10, forming a key connection in the central area.

From the perspective of isolated nodes, node 12 (“zdisease_12_”) has the lowest frequency of connection to other diseases, and the probability of being an independent disease or co-disease with other diseases is extremely low. The connection frequency of nodes 11 (“zdisease_11_”) and 13 (“zdisease_13_”) is moderately low, but there is still a certain correlation and may participate in some comorbid combinations as secondary diseases.

From a critical point of view “zdisease_1_” and “zdisease_14_” belong to a high-frequency co-occurring disease combination, and this combination may have common risk factors (such as lifestyle, environment, inheritance). “zdisease_10_” belongs to a disease bridging node, which may play a "hub" role in the comorbidity network and may become the key to prevention and intervention. “ Zdisease_12_”may be an isolated disease relative to the other 14 diseases, or it may be because its co-occurrence with other diseases is underestimated due to insufficient samples or diagnostic criteria. Overall, highly co-occurring diseases may require comprehensive management, while isolated diseases may require specific measures.

Further cluster analysis of the comorbidity patterns of these 15 chronic diseases can be carried out, and the results shown in Figure 3 can be obtained.

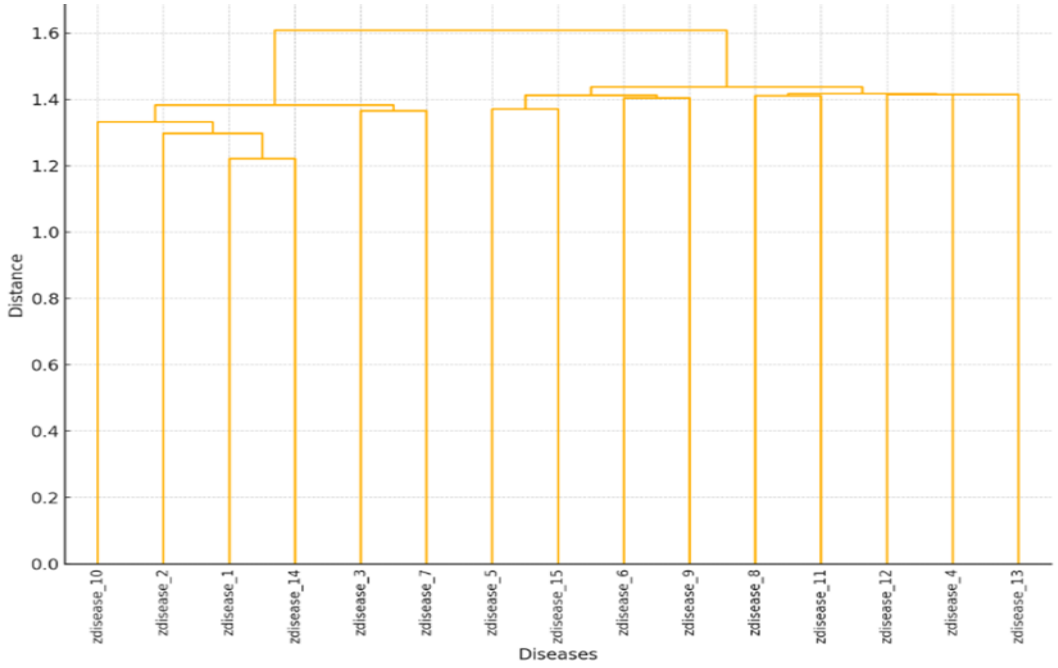


Figure 3: Hierarchical clustering of chronic diseases.

You can see more characteristics based on the hierarchical cluster diagram:

Zdisease_1 and zdisease_14 are clearly located in the same cluster branch, indicating that the two diseases have a high frequency of co-occurrence and may have a strong pathological or epidemiological relationship. At the same time, this group is also connected to branches of zdisease_10 and zdisease_5, suggesting that these diseases may form a core network with high co-occurrence. Zdisease_4 and zdisease_12 show independent branches, which may indicate that they have low co-occurrences with other diseases and have high independence.

The clustering results show that certain diseases form small subgroups (such as zdisease_2 and zdisease_3), and diseases within these subgroups may have clinical comorbidity tendencies. Zdisease_6 and zdisease_7 are not closely connected to the main branch, indicating that their co-occurrence characteristics are relatively special and may suggest some specific pathological processes or the influence of a single factor.

Zdisease_14 and zdisease_1 are in the central branch of the cluster diagram and are connected to multiple other disease branches, suggesting that they may play an important bridging role in the overall comorbidity network.

### 4.3. Negative Binomial Regression Model Analysis Results

After analysis using a negative binomial regression model, the goodness-of-fit indicator showed that the model fit well. The deviation ratio (Deviation/df = 0.744) and Pearson"s chi-square ratio (Pearson $\chi^2$/df = 0.568) were both less than 1, indicating that the model was in good agreement with the data. The information criteria (AIC = 25466.315 and BIC = 25610.018) provide a baseline for subsequent model comparisons. The Omnibus test was significant ($p < 0.001$), indicating that the variables included in the model significantly improved the interpretation of dependent variables. See Table 4. The calibration parameter of the negative binomial model is fixed at 1, indicating that the model is reasonable. See Table 5.

In terms of variable effects, the five factors Gender (ba001, p = 0.385), Marital Status (ba011, p = 0.604), Health Insurance (ba017, p = 0.760), Region (ba008, p = 0.457), and Depression (dc025, p = 0.924) had no significant impact on the number of comorbidities.

The Wald chi-square (63.344, $p < 0.001$) corresponding to Smoking (da047) was significant, indicating that smoking has a significant impact on the number of comorbidities. Estimates of specific parameters for smoking show: A comparison of still smoking (da047 =1) versus never smoking (da047 =3), parameter estimate B = -0.178, which indicates that the number of comorbidities in still smokers is reduced by $e^{-0.178} \approx 0.837$ times compared to never-smokers. In other words, the number of comorbidities in current smokers is approximately 16.3% lower than in never-smokers. Statistical significance p = 0.002 indicates that the result is statistically significant and smoking behavior does have a significant impact on the number of comorbidities. For the comparison of "quitting smoking (da047 =2)" and "never smoking (da047 =3)", the parameter estimate is B = 0.078. The number of comorbidities among smokers is slightly higher than that of "never smokers", but the B value is positive and not significant. Statistical significance p = 0.198 (> 0.05), indicating that there was no significant difference in the number of comorbidities between quitters and never-smokers.

Table 4: Goodness of fit[b].

| | value | Df | Statistics/degrees of freedom |
|---|---|---|---|
| Dispersion | 5136.878 | 6905 | .744 |
| Adjusted deviation | 5136.878 | 6905 | |
| Pearson chi-square | 3921.491 | 6905 | .568 |
| Adjusted Pearson Chi-Square | 3921.491 | 6905 | |
| Log likelihood a | -12712.157 | | |
| Akaike Information Guidelines (AIC) | 25466.315 | | |
| AICC Guidelines | 25466.449 | | |
| BIC Guidelines | 25610.018 | | |
| CAIC Guidelines | 25631.018 | | |

Notes: Dependent variable: comorbidity Count. Model: (intercept), gender (ba001, da047, da051), marriage (ba011), medical insurance (ba017), region (ba008, dc025, CCda033_0).

Alcohol (da051) corresponds to a significant Chi-square of Wald (54.310, $p < 0.001$), indicating that drinking alcohol has a significant impact on the number of comorbidities. Parameter estimates showed that da051 =1 and da051 =2 were significantly lower than da051 =3 ($p < 0.05$). Drinking variables are divided into three categories: da051 =1 (frequent drinking, current drinking frequency is high), da051 =2(occasional drinking, low drinking frequency), and da051 =3(never drinking, non-drinkers, baseline reference group). B =-0.237, $p < 0.001$ for regular drinkers (parameter da051 =1), compared with those who never drank, the number of comorbidities in regular drinkers was significantly lower. A negative B value means that the number of comorbidities among regular drinkers is $e^{-0.237} \approx 0.789$ times that of never drinkers, a decrease of approximately 21.1%. The B =-0.111, p = 0.031 for occasional drinkers (da051 =2), and the number of comorbidities among occasional drinkers was also significantly lower than those who never drank. A negative B value means that the number of comorbidities among occasional drinkers is $e^{-0.111} \approx 0.895$ times that of never drinkers, a decrease of approximately 10.5%.

The Wald chi-square (10.591, p = 0.001) of Physical Activity (CCda033_0_) was significant, indicating that physical activity had a significant impact on the number of comorbidities. Parameter estimates showed that increased physical activity (B = -0.025) reduced the number of comorbidities. Parameter B =-0.025, p = 0.001, the B value is negative, indicating that the greater the amount of physical activity, the lower the number of comorbidities. For every increase in physical activity, the number of comorbidities decreased by $e^{-0.025} \approx 0.975$ times. In other words, for every increase in physical activity, the number of comorbidities decreased by about 2.5%. See Table 5.

### 4.4. Model Evaluation

In this negative binomial regression model, the model is mainly evaluated from the following aspects to ensure that its fitting quality and prediction ability are scientifically verified.

#### 4.4.1. Overall Evaluation of the Model

The Omnibus test was used to compare whether the model with the predictor variable was significantly different from the model containing only the intercept. The results showed that the likelihood ratio chi-square value was 164.774, df = 20, and $p < 0.001$. The results of this test are significant, indicating that models containing predictive variables have significantly improved fitting compared to models only with intercept-only. This means that the independent variables (alcohol consumption, smoking, physical activity, etc.) do have an impact on the dependent variable (Comorbidity Count).

Look at the log-likelihood value as −12712.157. The larger the log-likelihood value (the smaller the absolute value), the better the model fits. AIC = 25466.315，BIC = 25610.018。These criteria are used to compare the goodness of fit of different models. The smaller the value, the better the model fits. In this model, AIC and BIC values are provided and can be used as a basis for comparison with other models.

#### 4.4.2. Goodness of Fit Evaluation

Deviance measures the degree to which model predictions deviate from actual observations. Values of dispersion/degrees of freedom are close to 1, indicating that the model fits well. If the value is significantly greater than 1, the model may be under-fitted; if it is significantly less than 1, the model may be over-fitted. The results showed dispersion = 5136.878, df = 6905, and dispersion/degree of freedom = 0.744.

Pearson"s chi-square can further test the goodness of fit of the model. The results showed Pearson"s chi-square/degrees of freedom = 0.568. A value of chi-square/degree of freedom is less than 1, indicating that the model fits well, but there may be a slight overfitting.

Table 5: Parameter estimation.

| Parameters | B | Standard error | 95% Wald confidence interval | | Hypothesis testing | | |
|---|---|---|---|---|---|---|---|
| | | | Lower limit | Upper limit | Wald chi-square | Df | Sig. |
| (intercept) | .355 | .5791 | -.780 | 1.490 | .376 | 1 | .540 |
| [ba001=1] | -.027 | .0306 | -.087 | .033 | .754 | 1 | .385 |
| [ba001=2] | 0a | . | . | . | . | . | . |
| [da047=1] | -.178 | .0588 | -.294 | -.063 | 9.216 | 1 | .002 |
| [da047=2] | .078 | .0610 | -.041 | .198 | 1.655 | 1 | .198 |
| [da047=3] | 0a | . | . | . | . | . | . |
| [da051=1] | -.237 | .0321 | -.300 | -.174 | 54.300 | 1 | .000 |
| [da051=2] | -.111 | .0517 | -.213 | -.010 | 4.649 | 1 | .031 |
| [da051=3] | 0a | . | . | . | . | . | . |
| [ba011=1] | -.061 | .2088 | -.470 | .348 | .086 | 1 | .770 |
| [ba011=2] | -.065 | .2143 | -.485 | .355 | .093 | 1 | .760 |
| [ba011=3] | -.325 | .3104 | -.934 | .283 | 1.099 | 1 | .295 |
| [ba011=4] | -.197 | .2485 | -.684 | .290 | .629 | 1 | .428 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| [ba011=5] | -.015 | .2125 | -.431 | .402 | .005 | 1 | .945 |
| [ba011=6] | 0a | . | . | . | . | . | . |
| [ba017=1] | .196 | .1322 | -.063 | .455 | 2.201 | 1 | .138 |
| [ba017=2] | .180 | .1331 | -.081 | .441 | 1.821 | 1 | .177 |
| [ba017=3] | .229 | .1446 | -.054 | .513 | 2.515 | 1 | .113 |
| [ba017=4] | .183 | .1253 | -.063 | .428 | 2.126 | 1 | .145 |
| [ba017=5] | .179 | .1898 | -.193 | .551 | .889 | 1 | .346 |
| [ba017=6] | 0a | . | . | . | . | . | . |
| [ba008=1] | .422 | .5227 | -.602 | 1.447 | .653 | 1 | .419 |
| [ba008=2] | .348 | .5235 | -.678 | 1.374 | .443 | 1 | .506 |
| [ba008=3] | .406 | .5221 | -.617 | 1.429 | .605 | 1 | .437 |
| [ba008=4] | 0a | . | . | . | . | . | . |
| dc025 | -1.197E-5 | .0001 | .000 | .000 | .009 | 1 | .924 |
| CCda033_0 | -.025 | .0077 | -.040 | -.010 | 10.591 | 1 | .001 |
| (scale) | 1b | | | | | | |
| (negative binomial) | 1 | | | | | | |

Dependent variable: Comorbidity Count. Model: (intercept), ba001, da047, da051, ba011, ba017, ba008, dc025, CCda033_0

## 5. Discussion

### 5.1. Discussion of the Main Findings

#### 5.1.1. The Main Patterns and Characteristics of Comorbidity of Chronic Diseases

Network structure characteristics: Most chronic diseases are concentrated in the central area of the network, showing a high degree of co-occurrence, which may be driven by common risk factors or pathological mechanisms. Zdisease_1_and zdisease_14_are high-frequency co-occurring diseases, and their importance is reflected in their high incidence and bridging effect, while zdisease_12_appears to be isolated diseases, which may be caused by low co-occurring frequency or specific sample characteristics.

Clustering relationship: zdisease_1_and zdisease_14_form a core network, forming closely connected branches with zdisease_10_and other diseases. This may indicate that these diseases share certain pathological mechanisms (e.g., inflammation, metabolic disorders). Zdisease_12_and zdisease_4_appear as independent branches, suggesting their unique etiological background.

Role of bridging node: zdisease_10_plays a "hub" role in the network. Its multi-connection characteristics suggest its key role in the disease process and are worth in-depth exploration in future research.

#### 5.1.2. Important Factors Affecting Comorbidity of Chronic Diseases

Current smoking is negatively correlated with the number of comorbidities, which may be explained by the fact that current smokers may have certain compensation mechanisms for other health behaviors (such as diet or exercise), which may temporarily reduce their number of comorbidities. It may also be that the number of comorbidities among current smokers has not yet accumulated significantly, and health problems caused by smoking will take a long time to fully manifest. The number of comorbidities among quitters is higher than that of never-smokers. This may be explained that quitters may have accumulated certain health damage during smoking, and it still

takes a long time to recover even after quitting smoking. Quitters may have other concomitant risk factors (such as higher age, higher baseline disease risk) that lead to higher comorbidities.

Drinking alcohol had a significant impact on the number of comorbidities. Frequent drinkers B =-0.237, $p < 0.001$. Compared with those who never drank, the number of comorbidities in frequent drinkers was significantly lower. Moderate drinking has lower comorbidities among regular and occasional drinkers than those who never drink. The possible explanation is the "protective effect" of moderate drinking. Moderate drinking (especially the polyphenol compounds in red wine) may be beneficial to cardiovascular health, thereby indirectly reducing the incidence of certain comorbidities. Regular drinkers may have a certain regularity and have a more stable amount of alcohol consumption, while occasional drinkers have lower amounts of alcohol consumption and have limited negative impact on health. Never drinkers may include some people who choose to abstain from alcohol for health reasons, and these people may themselves have a higher baseline risk of comorbidities. Drinking habits may be associated with diet, social activities, mental health, etc., which in turn affects the number of comorbidities.

The increase in physical activity was significantly negatively related to the number of common diseases. A possible explanation is the protective effect of physical activity: regular physical activity improves cardiopulmonary function, enhances immunity, reduces the risk of metabolic diseases (such as diabetes, high blood pressure), and thereby reduces the number of comorbidities. Physical activity may also relieve stress and improve mental health, which also has a positive effect on reducing the number of comorbidities. There is a positive correlation between activity and health behaviors. Individuals with high physical activity often accompany other health behaviors (such as healthy eating, not smoking). The combined effect of these behaviors may further reduce the number of comorbidities. Individual differences in activity, individuals with high physical activity may have better fitness or lower baseline health risks, which helps reduce comorbidities.

#### 5.1.3. Possible Mechanisms

Mechanisms of co-occurrence of diseases. Highly co-occurring diseases (such as zdisease_1_and zdisease_14_) may be driven by systemic mechanisms such as inflammation and metabolic disorders. These diseases may form a tight network through shared pathological processes or environmental risk factors.

The role of behavioral factors. Healthy behaviors (such as moderate drinking and regular exercise) may slow down the progression of chronic diseases by improving physiological functions and enhancing immunity. On the contrary, bad behaviors (such as smoking) may trigger mechanisms such as inflammation and immunosuppression, accelerating the development of comorbidities.

### 5.2. Practical Application Significance

#### 5.2.1. Enlightenment for Comprehensive Management of Chronic Diseases

Focus on managing high-risk disease combinations. Research results suggest that high-frequency co-occurring diseases (such as zdisease_1_and zdisease_14_) should be the focus of chronic disease management. These diseases may have common risk factors, and comprehensive interventions (such as controlling blood pressure and reducing inflammation) may be more effective.

Identify critical bridging nodes. Bridging diseases such as zdisease_10_play a pivotal role in the comorbid network, suggesting that these diseases may be key nodes in the disease transmission chain. Early intervention in bridging nodes may effectively cut off disease transmission paths, thereby reducing the overall disease burden.

Personalized management strategies. For isolated diseases (such as zdisease_12_), specific

interventions need to be developed. The management of these diseases may require a combination of personalized diagnosis and treatment and precision medical technology.

### 5.2.2. Healthy Aging Policy Recommendations

Promote healthy behaviors. Physical activity and moderate drinking have shown protective effects, suggesting that public health policies should further promote national sports programs and carry out healthy drinking education for different groups.

Support the recovery of smokers. It takes a long time for smokers to recover from health, so comprehensive support, including psychological counseling, health checkups and behavioral intervention, should be provided to individuals who quit smoking to help them better recover to health.

Focus on high-risk groups. The high risk of comorbidities among never-drinkers suggests attention to potential health issues, such as drinking alcohol due to existing chronic conditions. Such high-risk groups should be the focus of policy attention.

Establish an interdisciplinary intervention system. The complexity of comorbidity of chronic diseases requires interdisciplinary comprehensive intervention, combining multi-level strategies of medicine, psychology and public health to jointly address the challenges of comorbidity.

Optimize the chronic disease surveillance network. Based on the results of the comorbidity network, it is recommended to build a dynamic monitoring system for chronic diseases based on big data to track changes in disease associations and distribution in real time to provide scientific basis for policy formulation.

## 5.3. Research Limitations and Future Research Directions

This study is based on specific sample data and may not fully reflect the comorbidity patterns in the broader population. For example, different regions, age, and gender groups may have unique comorbidity characteristics, and these heterogeneity could not be fully included in the analysis. Although this study covered multiple health behaviors and demographic variables, other key influencing factors (such as genetic factors, environmental exposure, and access to medical resources) may be missed. Because cross-sectional data were used, this study only revealed the correlation of comorbidities of chronic diseases, and the causal relationship between health behaviors and comorbidities of chronic diseases could not be clarified.

Future research can introduce multi-center data with larger sample sizes and focus on the heterogeneity of epidemic patterns in different populations (such as the elderly, children, and specific occupational groups). At the same time, we explored the integrated analysis of genomics and environmental exposure data to reveal the deep mechanism of comorbidity formation. Through longitudinal data collection, we dynamically tracked the development process of chronic comorbidities to clarify the long-term impact of health behaviors on the risk of comorbidities and the causal chain between diseases. We applied complex network analysis methods to build a more accurate prediction model of comorbid diseases, studied the dynamic effects of key nodes in the network and bridging diseases on the process of comorbid diseases, and provides specific guidance for public health interventions. Based on comorbidity patterns and network characteristics, we developed personal-oriented health management tools, and combined artificial intelligence technology to achieve intelligent and refined comprehensive management of chronic diseases. Through field intervention experiments or policy simulations, we verified the actual impact of strategies such as health behavior promotion, key disease management, and bridging disease intervention on the burden of chronic diseases, and provided empirical basis for optimizing healthy aging policies.

This study not only reveals complex patterns of chronic comorbidities and their influencing factors,

but also proposes comprehensive management recommendations for high-risk disease combinations and health behaviors. Despite data and methodological limitations, this study provides important reference for chronic disease prevention and control and health policy optimization, and lays the foundation for future research.

## 6. Conclusion

This study systematically revealed the comorbidity patterns and network characteristics of chronic diseases through network analysis and hierarchical clustering methods. The results show that high-frequency co-occurring diseases (such as zdisease_1_and zdisease_14_) are concentrated in the center of the network, forming a strongly correlated core network. These diseases may be driven by common pathological mechanisms (such as inflammation, metabolic disorders) or risk factors (such as unhealthy lifestyles). Isolated diseases (such as zdisease_12_) may have high independence, suggesting unique etiology or sample characteristics. Negative binomial regression analysis showed that health behaviors (such as physical activity and moderate alcohol consumption) were significantly negatively associated with the number of common illnesses. For every unit increase in physical activity, the number of comorbidities decreased by approximately 2.5%; moderate drinkers (frequent or occasional drinkers) had a significantly lower risk of comorbidities than those who never drank. On the contrary, the impact of smoking behavior on the risk of comorbidities is more complex, and current smokers may have a lower risk of comorbidities because health compensation mechanisms or long-term damage has not yet emerged. Studies have found that bridging diseases such as zdisease_10_play an important pivotal role in the comorbid network and may connect multiple high-risk disease groups and are important entry points for comorbid prevention and control. The research results provide a scientific basis for comprehensive management of chronic diseases, emphasizing the need to prioritize intervention in combinations of highly co-occurring diseases, support the popularization of healthy behaviors, and implement personalized management strategies for different characteristic groups (such as smokers and patients with isolated diseases).

## References


*[1] Li, Y. N., & Wang, Y. Q. Prevalence and patterns of multimorbidity among elderly people in China (in Chinese). Chinese General Practice, 2021, 24(31), 3955-3962. https://doi.org/10.12114/j.issn.10079572.2021.00.295*

*[2] Feinstein, A. R. The pre-therapeutic classification of comorbidity in chronic disease. Journal of Chronic Diseases, 1970, 23(7), 455-468. https://doi.org/10.1016/0021-9681(70)90054-8*

*[3] Pan, Y., Liu, Z. H., Hu, Q. Q., et al. Patterns of coexistence of multiple chronic conditions among elderly Chinese (in Chinese). Chinese General Practice, 2023, 26(29), 3608-3615.*

*[4] He, L., Zhang, Y. F., Shen, X. C., et al. Trends in the prevalence of comorbidity among residents in Chinese mainland: A meta-analysis (in Chinese). Chinese General Practice, 2023, 26(29), 3599-3607. https://doi.org/10.12114/j.issn.1007-9572.2023.0217*

*[5] Yan, Z. Y., & Chen, D. F. Research progress on the patterns of multimorbidity (in Chinese). Chinese Journal of Disease Control and Prevention, 2023, 27(4), 471-475. https://doi.org/10.16462/j.cnki.zhjbkz.2023.04.017*

*[6] Xu, Z. J., Qian, Y., Yan, M., et al. The influencing factors and management strategies of multimorbidity based on syndemic theory (in Chinese). Chinese General Practice, 2024, 27(11), 1288-1295.*

*[7] Tang, S. S., & Jian, W. Y. The association between socioeconomic status and chronic disease: An empirical study based on working populations in Beijing and Shanghai (in Chinese). Chinese Health Policy Research, 2012, 1(1), 51-55.*

*[8] Wu, D. M. A study on the correlation between multimorbidity and lifestyle behaviors among middle-aged and elderly people in China and the conception of nursing mode [D] (in Chinese). Hubei University of Medicine. 2021.*

*[9] Hu, Q. Q., Zhou, T., Liu, Z. H., Shi, W. L., & Wang, L. Y. Research progress on patterns and management strategies of comorbidity in elderly patients with chronic diseases(in Chinese). Chinese Journal of General Practitioners, 2023, 22(7), 754-758. https://doi.org/10.3760/cma.j.cn114798-20230411-00312*

*[10] Wu, X. M., Jiang, X. B., Zhao, W. J., Jia, S. Y., Miao, Y. D., Cao, Z. H., & Wu, J. Impact of social and economic status on the prevalence of chronic diseases and comorbidity among middle-aged and elderly people in Henan Province(in Chinese). Medicine and Society, 2024 (10), 42-48. doi:10.13723/j.yxysh.2024.10.007.*

*[11] Xin B., Wu Y. X., Zhang D., He Y. X., Yang S., Li M. C. & Jiang W.H. Research progress on cognitive dysfunction in patients with chronic comorbidity. China General Practice Medicine, 2025 (02), 143-148+158.*
*[12] Mu, X. Y., Zhao, M. J., Ma, T. P., Xu, J., Shi, Y., Li, S. Q., & Su, Q. L. The effect of social support on multimorbidity and related outcomes among middle-aged and older adults in China. Journal of Sichuan University (Medical Science Edition), 2022, 53(5), 872-879. https://doi.org/10.12182/20220960507 ; PMID: 36224691; PMCID: PMC10408810*