scientific reports

Check for updates

OPEN

Exploring patterns of multimorbidity in South Korea using exploratory factor analysis and non negative matrix factorization

Yeonjae Kim[1], Samina Park[2], Yun Mi Choi[3], Byung-Ho Yoon[4], Su Hyun Kim[5], Jin Park[6], Hyun Jin Oh[7], Yaeji Lim[8], Jungkyun Lee[8] & Bomi Park[1]✉

The increasing prevalence of multimorbidity and the co-occurrence of multiple chronic diseases presents a measurable challenge to public health, impacting healthcare strategies and planning. This study aimed to explore disease patterns and temporal clustering using data from South Korea's National Health Insurance Service, spanning 2002–2019. The dataset included approximately 1 million individuals, focusing on those with at least two chronic diseases while excluding individuals who died within five years of follow-up. We analyzed 126 non-communicable diseases, considering only those with a prevalence above 1%, and applied a wash-out period to determine incidence. Exploratory factor analysis (EFA) and non-negative matrix factorization (NMF) were used to identify disease clustering over time. Participants were divided into four groups: men and women in their 50 s and 60 s. EFA identified five patterns in men in their 50 s and seven in their 60 s, while four patterns emerged in women in their 50 s and five in their 60 s. NMF identified 10 clusters for men in their 50 s, 15 in their 60 s, and 16 clusters for women in both age groups. Our study confirms established comorbidity patterns and reveals previously unrecognized clusters, providing data-driven insights into multimorbidity mechanisms and supporting evidence-based healthcare strategies.

Keywords Multimorbidity, Chronic disease, EFA, NMF, Longitudinal

Abbreviations

EFA	Exploratory factor analysis
NHIS	National health insurance service
ICD-10	International classification of disease 10th revision
KMO	Kaiser–Meyer–Olkin
NMF	Non-negative matrix factorization
LBP	Low back pain
BPH	Benign prostatic hyperplasia
OA	Osteoarthritis
TTH	Tension-type headaches
AFF	Atrial fibrillation and flutter
COPD	Chronic obstructive pulmonary disease

[1]Department of Preventive Medicine, College of Medicine, Chung-Ang University, Seoul, Korea. [2]Department of Thoracic and Cardiovascular Surgery, Seoul National University Hospital, Seoul National University College of Medicine, Seoul, Korea. [3]Department of Internal Medicine, Hallym University Dongtan Sacred Heart Hospital, Hallym University College of Medicine, Hwaseong, Korea. [4]Department of Orthopedic Surgery, College of Medicine, Ewha Womans University, Mokdong Hospital, Seoul, Korea. [5]Department of Internal Medicine, Chung-Ang University Gwangmyeong Hospital, Chung-Ang University College of Medicine, Gwangmyeong, Korea. [6]Department of Neurology, Asan Medical Center, University of Ulsan College of Medicine, Seoul, Republic of Korea. [7]Division of Gastroenterology, Department of Internal Medicine, Center for Cancer Prevention and Detection, National Cancer Center, Goyang, Korea. [8]Department of Applied Statistics, Chung-Ang University, Seoul, Korea. ✉email: bpark@cau.ac.kr

IHD	Ischemic heart disease
CKD	Chronic kidney disease
PUD	Peptic ulcer disease
GERD	Gastroesophageal reflux disease
ILD	Interstitial lung disease
RA	Rheumatoid arthritis
UTI	Tubulointerstitial nephritis, pyelonephritis, and urinary tract infections
PVD	Peripheral vascular disease
NSAIDs	Nonsteroidal anti-inflammatory drugs
H. pylori	Helicobacter pylori

Multimorbidity, defined as the concurrent presence of two or more diseases in an individual, poses a measurable challenge to public health systems globally. The prevalence of chronic diseases is escalating due to lifestyle changes and environmental factors, while advances in medical technology have reduced mortality rates from these conditions[1, 2]. Furthermore, with life expectancy increasing over time—exemplified by an average life expectancy of 83.6 years in Korea in 2021[3]—more individuals are living with multiple chronic diseases. With Korea's rapidly aging population, multimorbidity prevalence is projected to rise further, impacting healthcare demands and policy planning.

Multimorbidity significantly influences health outcomes by increasing the use of healthcare and social care services and is associated with disability, diminished quality of life, and higher healthcare costs, underscoring its implications for individuals and healthcare systems[4–9]. The traditional healthcare system, which predominantly focuses on single diseases, fails to address the complexities of multimorbidity, leading to treatment burdens, economic strains, therapeutic conflicts, and poorer prognoses[10–12]. Therefore, it is crucial to shift from single-disease-centric approaches to a more holistic consideration of multiple diseases in medical services and public health policies to reduce the physical and economic burdens of multimorbidity and improve outcomes. Understanding comorbidity patterns is essential for developing effective strategies to prevent and manage multimorbidity.

Numerous studies have aimed to identify patterns of multimorbidity using various methodologies. Cluster analysis has been widely used to identify patterns in multimorbidity, grouping individuals with similar characteristics or disease profiles[13–19]. Latent class analysis is another common method, identifying distinct groups of individuals with similar combinations of multiple chronic conditions[20–24]. Exploratory factor analysis (EFA) also falls into this category, uncovering underlying relationships between variables to group individuals with high probabilities of having potential comorbid chronic conditions[25–28]. Additionally, principal component analysis and hidden Markov models have been utilized in some studies to identify patterns of co-occurring diseases[29, 30]. Network analysis has been employed to examine comorbidity between conditions, identifying diseases that frequently appear together[13, 31–33].

However, most previous studies on multimorbidity patterns have considered only a limited number of conditions, examined a small number of co-occurring clusters, or analyzed data cross-sectionally without temporal phenotyping, thereby limiting the understanding of chronic disease multimorbidity patterns over time. Therefore, this study aims to consider all comorbidities simultaneously and address the temporal aspects of multimorbidity using large-scale longitudinal data. We analyze disease onset through cohort follow-up using various statistical methodologies, ranging from traditional statistical approaches, such as EFA for identifying patterns, to machine learning techniques, such as non-negative matrix factorization (NMF), for uncovering longitudinal patterns.

Methods

Data and study population

This study used retrospective cohort data from the National Health Insurance Service (NHIS) in South Korea, covering the period from 2002 to 2019. The NHIS, a comprehensive health insurance system, provides coverage for the entire South Korean population. Due to its universal coverage, the NHIS maintains healthcare utilization information for nearly the entire population through its claims data. For research purposes, the NHIS developed a sample cohort database containing anonymized demographic and clinical information on all public healthcare system users in South Korea[34]. To create this sample, the NHIS extracted data from about one million individuals by stratifying 2% of the national population eligible for health insurance and medical benefits in 2006. This stratification accounted for variables such as sex, age, income level, and residential area. From this sample cohort, we extracted data on sex, age, diagnosis codes, hospitalizations, and outpatient visits for adults aged 50 to 69 as of 2002. Focusing on multimorbidity patterns, the study included only subjects with at least two chronic diseases and excluded individuals who died within five years of follow-up.

The Institutional Review Board of Chung-Ang University reviewed the study protocol and determined it meets the requirements eligible for an exemption from ethical approval. Consequently, Informed consent was waived by the approval committee due to the use of anonymized data, in compliance with the Act on Bioethics and Safety. This decision is recorded under approval number 1041078–202,112-HR-336–01.

Disease definitions

This study followed the disease classification system and definitions for incidence and prevalence proposed by Yoon et al.[35]. Chronic diseases such as hypertension and dyslipidemia were excluded from the analysis because they are well-known precursor conditions for other chronic diseases[36–38] and are classified as risk factors rather than standalone diseases. Our analysis included 126 non-communicable diseases (Table S1). To determine the annual prevalence of each disease, we used the International Classification of Disease 10th revision (ICD-10)

codes along with data on the number of hospitalizations and outpatient visits (ranging from 1–5 times) within a given year. Disease incidence was identified by implementing a washout period of one to five years, depending on the specific disease. This approach facilitated precise tracking of annual prevalence and incidence for 126 diseases from 2007 to 2019. To ensure clinical significance and avoid misinterpretation of relationships, only diseases with a prevalence of 1% or more in each age and sex group were included in the analyses.

Statistical analysis

Analyses were stratified into four groups based on age (50–59 and 60–69 years) and sex (male and female). Descriptive statistics characterized participants' demographic and clinical features.

EFA was conducted using a correlation matrix to identify patterns of multimorbidity. Individuals with a prevalent condition diagnosed between 2002 and 2019 were included in the analysis based on disease prevalence data. Given the dichotomous nature of the variables, a tetrachoric correlation matrix was employed, as it is more appropriate for binary variables than Pearson's correlation, which assumes continuous data, or polychoric correlation, typically utilized for ordinal variables. Sample adequacy for EFA was assessed using the Kaiser–Meyer–Olkin (KMO) measure, with groups exhibiting a KMO below 0.6 excluded from the analysis. Bartlett's test of sphericity further confirmed the suitability of the correlation matrix. KMO values exceeded 0.6 for all subgroups, and Bartlett's test yielded p-values < 0.05, confirming the suitability of each subgroup for EFA. The optimal number of factors was identified using a scree plot analysis, adhering to the "elbow" criterion[39,40]. Oblique rotation (oblimin) was employed to facilitate the interpretability of the factors. Oblique rotation, unlike an orthogonal rotation, better captures the complex interrelationships among chronic diseases, making it more suitable for identifying multimorbidity patterns. Each pattern was defined by the factor loadings, which represent the strength of the association between a chronic disease and a latent factor. To identify clinically significant disease patterns, we excluded diseases with weak associations by establishing a factor-loading threshold, incorporating only those with a factor loading of ≥ 0.3 due to the adequate sample size of the study.

Next, we used non-negative matrix factorization (NMF), a machine learning technique, to analyze the temporal phenotyping of diseases using longitudinal datasets. NMF is a robust method for reducing the dimensionality of large datasets and imparting meaningful interpretations to matrices of non-negative data, widely recognized for its applications in data clustering[41]. Unlike EFA, which used disease prevalence data, NMF used disease incidence data. With a washout period between 1 and 5 years, disease occurrence was identified starting at age 55 for individuals in their 50 s and at age 65 for those in their 60 s. Thus, disease data from 2007 to 2019 were analyzed. The initial disease matrix was constructed as a disease × time matrix for each patient, with time represented by age. This matrix was standardized to ensure all patients had the same number of rows, from the minimum to the maximum age of all patients between 2007 and 2019. Before constructing the disease × time matrix, a wash-out period was applied to ensure each disease was recorded as its first occurrence at the specified age. Missing disease information for a given patient age was interpreted as absence of disease. The age range spanned 55 to 76 years for participants in their 50 s and 65 to 86 years for those in their 60 s. Variability in the incidence of disease can introduce a bias toward diseases that occur more frequently. To mitigate this, we applied the Disease Frequency-Inverse Patient Frequency adjustment[42], which reduces clustering bias around high-incidence diseases and ensures a balanced representation across conditions. Subsequently, a Gaussian kernel was applied to smooth the data from the time of medical facility visits and diagnoses. This step addresses the lack of temporal regularization in NMF and accounts for potential discrepancies between the recorded diagnosis time and the actual disease onset, which might have occurred years earlier. The smoothing strength was defined by using the standard deviation (σ). The optimal σ and appropriate number of clusters were determined using cophenetic correlation analysis. We evaluated σ values ranging from 1 to 5 and cluster numbers from 2 to 20. Optimization was performed separately for each sex and age group. The σ value and number of clusters yielding the highest cophenetic correlation were selected. Optimization yielded $\sigma = 1$ and 10 clusters for men in their 50 s, $\sigma = 2$ and 15 clusters for men in their 60 s, $\sigma = 1$ and 16 clusters for women in their 50 s, and $\sigma = 3$ and 16 clusters for women in their 60 s (see Figure S1). Data were smoothed using the selected σ values, and NMF was performed with the corresponding number of clusters. Smoothed patient matrices were then consolidated into a single matrix representing the final disease condition matrix. This matrix was then decomposed into matrix A and matrix B. Matrix A consists of the cluster and disease matrix, providing information about the diseases in each cluster, while matrix B comprises the patient's age and cluster matrix, detailing the time of cluster occurrence for each patient. Matrix A was used to identify the top three diseases with the highest weights in each cluster, with weights adjusted based on the highest weight in each cluster. Using matrix B, we selected the age of the patient with the maximum time weight in each cluster. We then plotted the mean and median distributions of disease onset ages within clusters. Since the data did not meet the assumptions for parametric analysis, differences in the order of onset between diseases were analyzed using the Kruskal–Wallis test, with statistical significance set at a p-value threshold of 0.05.

EFA was conducted using R version 4.2.2 (R Foundation for Statistical Computing, Vienna, Austria), while NMF was analyzed using R and Python (version 3.9.16). Cophenetic correlation analysis was performed using the Python NumPy package, and Gaussian kernel smoothing was implemented using the R plyr package. NMF was performed using the nimfa package, which utilizes a multiplicative update algorithm based on Kullback–Leibler divergence and includes enhancements to prevent numerical underflow[43,44].

Results

Baseline characteristics of study participants

A total of 142,505 individuals aged 50 to 69 years with multimorbidity (defined as the presence of two or more chronic diseases) were analyzed during the retrospective study period from 2002 to 2019. The cohort included 65,330 men (age 50–59: 37,833; age 60–69: 27,497) and 77,175 women (age 50–59: 41,161; age 60–69: 36,014)

(Table 1). On average, individuals in their 50 s had 5.34 chronic conditions (men: 5.16; women: 5.50), while those in their 60 s had 6.13 (men: 6.02; women: 6.21). An examination of prevalent conditions by sex and age (Table S2) showed that low back pain (LBP) was the most prevalent condition across all sexes and age groups, with a prevalence rate exceeding 50% in females. In men, benign prostatic hyperplasia (BPH) had the second-highest prevalence, at 33.4% in their 50 s and 41.1% in their 60 s. In women, osteoarthritis (OA) had the second-highest prevalence, at 53.9% in their 50 s and 62.4% in their 60 s. Gastrointestinal diseases were also highly prevalent, except among men in their 60 s. Regarding comorbidity frequency, BPH and LBP were the most common in males, whereas OA and LBP were predominant in females. Overall prevalence is higher in women than in men, and for OA in particular, women have approximately twice the prevalence of men in both age groups. Also, there was a notable co-occurrence of disease pairs among conditions with high individual prevalence rates.

EFA

Table 2 summarizes the disease clusters identified through EFA, stratified by sex and age group. Diseases are sorted in order of the highest factor loading in each factor. Table S3-S6 present additional EFA results, including explained variance and cumulative variance. All diseases were encompassed within a singular factor.

Men in their 50 s

Among men, five distinct patterns were identified in the 50 s age group.

- Factor 1: Gastrointestinal diseases, musculoskeletal diseases, migraine, tension-type headache (TTH), BPH, chronic obstructive pulmonary disease (COPD)
- Factor 2: Skin diseases
- Factor 3: Diabetes mellitus and chronic kidney disease (CKD) due to diabetes
- Factor 4: Stroke and atrial fibrillation and flutter (AFF)
- Factor 5: Cirrhosis of the liver, liver cancer

Among all groups, men in their 50 s are the only demographic in which COPD is linked to the diseases grouped under Factor 1.

Men in their 60 s Seven patterns were identified in men in their 60 s (Table 2).

- Factor 1: gastrointestinal diseases, musculoskeletal diseases, BPH, migraine, TTH, cataracts
- Factor 2: Skin diseases
- Factor 3: Diabetes mellitus and chronic kidney disease (CKD) due to diabetes
- Factor 4: Decubitus ulcer, stroke, Parkinson's disease, tubulointerstitial nephritis, pyelonephritis, and urinary tract infections (UTI)
- Factor 5: Liver diseases
- Factor 6: Pancreas diseases
- Factor 7: Respiratory diseases

Factors 1, 2, 3, and 5 exhibited a similar pattern to that observed in the male 50 s age group. However, a notable difference was the absence of COPD from Factor 1; instead, respiratory diseases, including asthma, formed a distinct cluster (Factor 7). In Factor 4, strokes were linked with Parkinson's disease, decubitus ulcers, and urinary tract infections. Factor 6, comprising pancreatic diseases, is also a pattern of disease combinations that is only seen in the male 60 s age group.

			Total	Male	Female
			(n = 142,505)	(n = 65,330)	(n = 77,175)
Age[a]		50–59	78,994 (57.0)	37,833 (59.6)	41,161 (54.7)
		60–69	63,511 (43.0)	27,497 (40.4)	36,014 (45.3)
No. of diagnoses[a]	Age 50–59	2	11,581 (14.7)	6324 (16.7)	5257 (12.8)
		3	12,345 (15.6)	6342 (16.8)	6003 (14.6)
		4	12,111 (15.3)	5943 (15.7)	6168 (15.0)
		≥5	42,957 (54.4)	19,224 (50.8)	23,733 (57.7)
	Age 60–69	2	6157 (9.7)	3091 (11.2)	3066 (8.5)
		3	7574 (11.9)	3535 (12.9)	4039 (11.2)
		4	8342 (13.1)	3736 (13.6)	4606 (12.8)
		≥5	41,438 (65.2)	17,135 (62.3)	24,303 (67.5)
Mean no. of diagnoses	Age 50–59		5.34	5.16	5.50
	Age 60–69		6.13	6.02	6.21
No. of diagnoses included	Age 50–59		50	49	45
	Age 60–69		54	50	56

Table 1. Characteristics of patients with multimorbidity. [a] N (%).

	Male						Female					
	Age 50 to 59	Age 60 to 69					Age 50 to 59			Age 60 to 69		
Factor	Disease	FL	pr	Disease	FL	pr	Disease	FL	pr	Disease	FL	pr
1	GERD	0.712	21.7	GERD	0.764	20.9	GERD	0.598	28.3	GERD	0.727	26.0
	Gastritis & duodenitis	0.699	23.4	Gastritis & duodenitis	0.660	28.6	Gastritis & duodenitis	0.536	33.2	Gastritis & duodenitis	0.657	36.1
	PUD	0.556	21.6	PUD	0.606	23.3	PUD	0.514	22.7	PUD	0.645	24.3
	LBP	0.461	37.1	LBP	0.453	45.2	Paralytic ileus & intestinal obstruction without hernia	0.335	1.2	TTH	0.362	6.4
	TTH	0.449	2.9	Neck pain	0.413	12.1	TTH	0.334	5.7			
	BPH	0.417	33.4	BPH	0.412	41.1	Pancreatitis	0.329	1.5			
	Neck pain	0.412	14.0	OA	0.384	31.5						
	Migraine	0.369	1.3	Migraine	0.381	1.9						
	OA	0.362	25.0	TTH	0.378	3.5						
	COPD	0.326	21.7	Cataracts	0.323	39.7						
2	Eczema	0.828	17.4	Eczema	0.786	19.8	Eczema	0.696	16.3	Eczema	0.666	17.3
	Psoriasis	0.568	1.6	Pruritus	0.581	4.2	Pruritus	0.555	2.6	Urticaria	0.534	12.5
	Urticaria	0.498	10.5	Urticaria	0.575	11.2	Urticaria	0.540	12.6	Pruritus	0.531	3.8
	Pruritus	0.473	3.0	Psoriasis	0.536	1.6				Abscess impetigo & other bacterial skin diseases	0.396	10.5
				Fungal skin diseases	0.365	10.2				Fungal skin disease	0.369	10.1
				Abscess impetigo & other bacterial skin diseases	0.317	11.6				OA	0.369	62.4
										Cellulitis	0.324	5.2
										LBP	0.322	65.0
										Viral skin diseases	0.306	2.6
3	CKD due to diabetes	1.024	1.0	CKD due to diabetes	0.997	1.4	OA	0.564	53.9	CKD due to diabetes	1.012	1.3
	Diabetes mellitus	0.633	25.3	Diabetes mellitus	0.652	26.0	LBP	0.532	56.2	Diabetes mellitus	0.681	27.7
							Neck pain	0.375	18.0			
							RA	0.300	6.6			
4	Ischemic stroke	0.704	13.0	Decubitus ulcer	0.687	1.6	Ischemic stroke	0.629	11.6	Decubitus ulcer	0.617	1.6
	Hemorrhagic & other non-ischemic stroke	0.669	2.5	Ischemic stroke	0.391	21.6	Hemorrhagic & other non-ischemic stroke	0.598	2.1	Ischemic stroke	0.374	21.2
	AFF	0.310	2.9	Parkinson's disease	0.384	3.2	Parkinson's disease	0.312	1.6	Hemorrhagic & other non-ischemic stroke	0.358	3.6
				Hemorrhagic & other non-ischemic stroke	0.358	3.7				Parkinson's disease	0.351	3.9
				UTI	0.344	4.9						
5	Cirrhosis of the liver	1.010	10.7	Liver cancer	1.037	3.0				AFF	0.603	3.7
	Liver cancer	0.509	2.6	Cirrhosis of the liver	0.522	8.7				Hypertensive heart disease	0.480	13.5
				Gallbladder & biliary tract cancer	0.505	1.3				IHD	0.399	16.2
										Gout	0.397	1.0
										CKD due to hypertension	0.332	2.3

Continued

	Male						Female					
	Age 50 to 59			Age 60 to 69			Age 50 to 59			Age 60 to 69		
Factor	Disease	FL	pr	Disease	FL	pr	Disease	FL	pr	Disease	FL	pr
6				Pancreatic cancer	1.050	1.0						
				Pancreatitis	0.395	2.0						
7				COPD	0.686	29.7						
				Asthma	0.655	14.9						

Table 2. Patterns of multimorbidity in EFA results stratified by age group and sex. Only diseases with factor loadings greater than 0.3 are included in the table. FL: factor loadings; pr: prevalence.

Women in their 50 s In women, four patterns were identified in their 50 s (Table 2).

- Factor 1: Gastrointestinal diseases, TTH
- Factor 2: Skin diseases
- Factor 3: Musculoskeletal diseases
- Factor 4: Stroke, Parkinson's disease

Gastrointestinal diseases and TTH constituted Factor 1 in women in their 50 s. Factor 3 consisted solely of musculoskeletal diseases, while Factor 2, similar to the pattern observed in men, comprised skin diseases. Lastly, Factor 4 consisted of stroke and Parkinson's disease.

Women in their 60 s Five patterns were identified in women in their 60 s (Table 2).

- Factor 1: Gastrointestinal diseases, TTH, pancreatitis
- Factor 2: Skin diseases, musculoskeletal diseases
- Factor 3: Diabetes mellitus, CKD due to diabetes
- Factor 4: Decubitus ulcer, stroke, Parkinson's disease
- Factor 5: AFF, hypertensive heart disease, ischemic heart disease (IHD), Gout, CKD due to hypertension

While Factor 1 showed a similar pattern to women in their 50 s, Factor 2 showed a new combination of skin and musculoskeletal diseases. Factor 3 remained consistent across all male age groups, while Factor 4 was similar to the combination observed in men in their 60 s, and Factor 6 presented a unique combination found exclusively in women in their 60 s.

NMF

Table 3 summarizes disease clusters and their associated weights, stratified by sex and age group, derived from NMF analysis.

Men in their 50 s The analysis revealed ten disease clusters for men in their 50s (Table 3) (Fig. 1). Diseases within each cluster were ordered by weight.

- Cluster 1: IHD, diabetes mellitus, hypertensive heart disease
- Cluster 2: musculoskeletal diseases, cirrhosis of the liver
- Cluster 3: PUD, hypertensive heart disease, LBP
- Cluster 5: Skin diseases, diabetes mellitus
- Cluster 4 and 7: Skin diseases
- Cluster 6: Eye diseases
- Cluster 8: Respiratory diseases, BPH
- Cluster 9: Stroke, diabetes mellitus
- Cluster 10: Gastrointestinal diseases, Gall bladder and bile duct disease

Clusters 4 and 7, both comprising skin diseases, and cluster 10, encompassing gastrointestinal and gallbladder/bile duct diseases, were also identified in the EFA results for men in their 50 s. Conversely, several other disease combinations were observed only in the NMF analysis.

Men in their 60 s The analysis identified 16 distinct clusters among men in their 60 s (Table 3) (Fig. 2).

- Cluster 1: Gastrointestinal diseases
- Cluster 2: Gastrointestinal diseases, TTH
- Cluster 3: Skin and eye diseases
- Cluster 4: Hypertensive heart disease, TTH, CKD due to hypertension
- Cluster 5 and 7: Skin diseases
- Cluster 6: IHD, AFF, BPH
- Cluster 8: Alzheimer's disease, Parkinson's disease, TTH
- Cluster 9: Stroke, UTI
- Cluster 10: Respiratory diseases, LBP
- Cluster 11: Liver diseases, gallbladder and biliary tract cancer
- Cluster 12: Musculoskeletal diseases
- Cluster 13: Lung and respiratory diseases
- Cluster 14: diabetes, OA, cellulitis
- Cluster 15: Eye diseases

Clusters 2, 3, 5, 7, 9, 10, 11, and 12, identified through EFA, were consistently reproduced in the NMF analysis, while the remaining clusters were unique to NMF.

Women in their 50 s Among women in their 50 s, the study identified 16 different clusters (Table 3) (Fig. 3).

- Cluster 1: Gastrointestinal diseases, TTH
- Cluster 2: Asthma, OA, diabetes
- Cluster 3: PUD, bacterial skin disease, OA

	Male				Female			
	Age 50 to 59		Age 60 to 69		Age 50 to 59		Age 60 to 69	
Pattern	Disease	Weight	Disease	Weight	Disease	Weight	Disease	Weight
1	IHD	1.000	PUD	1.000	GERD	1.000	AD & other dementias	1.000
	Diabetes mellitus	0.094	Gall bladder & bile duct disease	0.538	Gastritis & duodenitis	0.930	Parkinson's disease	0.071
	Hypertensive heart disease	0.087	Paralytic ileus & intestinal obstruction without hernia	0.077	TTH	0.404	PVD	0.054
2	Neck pain	1.000	GERD	1.000	Asthma	1.000	Asthma	1.000
	OA	0.500	Gastritis & duodenitis	0.602	OA	0.030	COPD	0.500
	Cirrhosis of the liver	0.473	TTH	0.115	Diabetes mellitus	0.023	OA	0.044
3	PUD	1.000	Fungal skin diseases	1.000	PUD	1.000	IHD	1.000
	Hypertensive heart disease	0.035	Viral skin diseases	0.037	Abscess impetigo & other bacterial skin diseases	0.602	Ischemic stroke	0.823
	LBP	0.035	Cataracts	0.032	OA	0.047	AFF	0.191
4	Eczema	1.000	Hypertensive heart disease	1.000	Ischemic stroke	1.000	Urticaria	1.000
	Pruritus	0.077	TTH	0.033	Diabetes mellitus	0.478	Pruritus	0.055
	Psoriasis	0.042	CKD due to hypertension	0.028	Hemorrhagic & other non-ischemic stroke	0.062	Gastritis & duodenitis	0.032
5	Abscess impetigo & other bacterial skin diseases	1.000	Urticaria	1.000	Fungal skin diseases	1.000	Abscess impetigo & other bacterial skin diseases	1.000
	Cellulitis	0.112	Eczema	0.879	Cellulitis	0.048	Cellulitis	0.194
	Diabetes mellitus	0.085	Pruritus	0.317	OA	0.028	Viral skin diseases	0.046
6	Cataracts	1.000	IHD	1.000	COPD	1.000	PUD	1.000
	Glaucoma	0.670	AFF	0.081	OA	0.038	GERD	0.704
	Refraction & accommodation disorders	0.241	BPH	0.033	LBP	0.038	Gastritis & duodenitis	0.398
7	Fungal skin diseases	1.000	Abscess impetigo & other bacterial skin diseases	1.000	RA	1.000	Gall bladder & bile duct disease	1.000
	Urticaria	0.939	Cellulitis	0.048	Cellulitis	0.066	Cirrhosis of the liver	0.119
	Cellulitis	0.058	Viral skin diseases	0.038	OA	0.048	Pancreatitis	0.086
8	Asthma	1.000	AD & other dementias	1.000	Urticaria	1.000	UTI	1.000
	COPD	0.984	Parkinson's disease	0.155	Pruritus	0.048	Parkinson's disease	0.084
	BPH	0.102	TTH	0.047	OA	0.024	Urolithiasis	0.072
9	Ischemic stroke	1.000	Ischemic stroke	1.000	UTI	1.000	TTH	1.000
	Hemorrhagic & other non-ischemic stroke	0.088	UTI	0.239	Urolithiasis	0.067	Migraine	0.248
	Diabetes mellitus	0.071	Hemorrhagic & other non-ischemic stroke	0.180	Urinary incontinence	0.062	PVD	0.080
10	Gastritis & duodenitis	1.000	Asthma	1.000	Gall bladder & bile duct disease	1.000	Hypertensive heart disease	1.000
	GERD	0.932	COPD	0.309	Cirrhosis of the liver	0.960	AFF	0.039
	Gall bladder & bile duct disease	0.257	LBP	0.036	Pancreatitis	0.073	Parkinson's disease	0.021
11			Cirrhosis of the liver	1.000	Eczema	1.000	Neck pain	1.000
			Liver cancer	0.260	Pruritus	0.066	LBP	0.044
			Gallbladder & biliary tract cancer	0.090	Viral skin diseases	0.048	Gastritis & duodenitis	0.040
12			Neck pain	1.000	Neck pain	1.000	Diabetes mellitus	1.000
			LBP	0.075	LBP	0.064	Cirrhosis of the liver	0.119
			OA	0.047	OA	0.041	Cataracts	0.070
13			Trachea bronchus & lung cancers	1.000	Hypertensive heart disease	1.000	Eczema	1.000
			COPD	0.074	Hemorrhoid	0.024	Pruritus	0.170
			ILD & pulmonary sarcoidosis	0.037	OA	0.024	Viral skin diseases	0.040
14			Diabetes mellitus	1.000	IHD	1.000	Fungal skin diseases	1.000
			OA	0.621	AFF	0.030	PVD	0.045
			Cellulitis	0.520	LBP	0.023	Refraction & accommodation disorders	0.035
15			Glaucoma	1.000	Glaucoma	1.000	RA	1.000
			Cataracts	0.198	Cataracts	0.561	PVD	0.070
			Refraction & accommodation disorders	0.160	Refraction & accommodation disorders	0.294	OA	0.037

Continued

Pattern	Male				Female			
	Age 50 to 59		Age 60 to 69		Age 50 to 59		Age 60 to 69	
	Disease	Weight	Disease	Weight	Disease	Weight	Disease	Weight
16					Varicose veins of lower extremities	1.000	Glaucoma	1.000
					OA	0.028	Refraction & accommodation disorders	0.211
					PVD	0.025	Cataracts	0.136

Table 3. Multimorbidity clusters in NMF stratified by age group and sex.

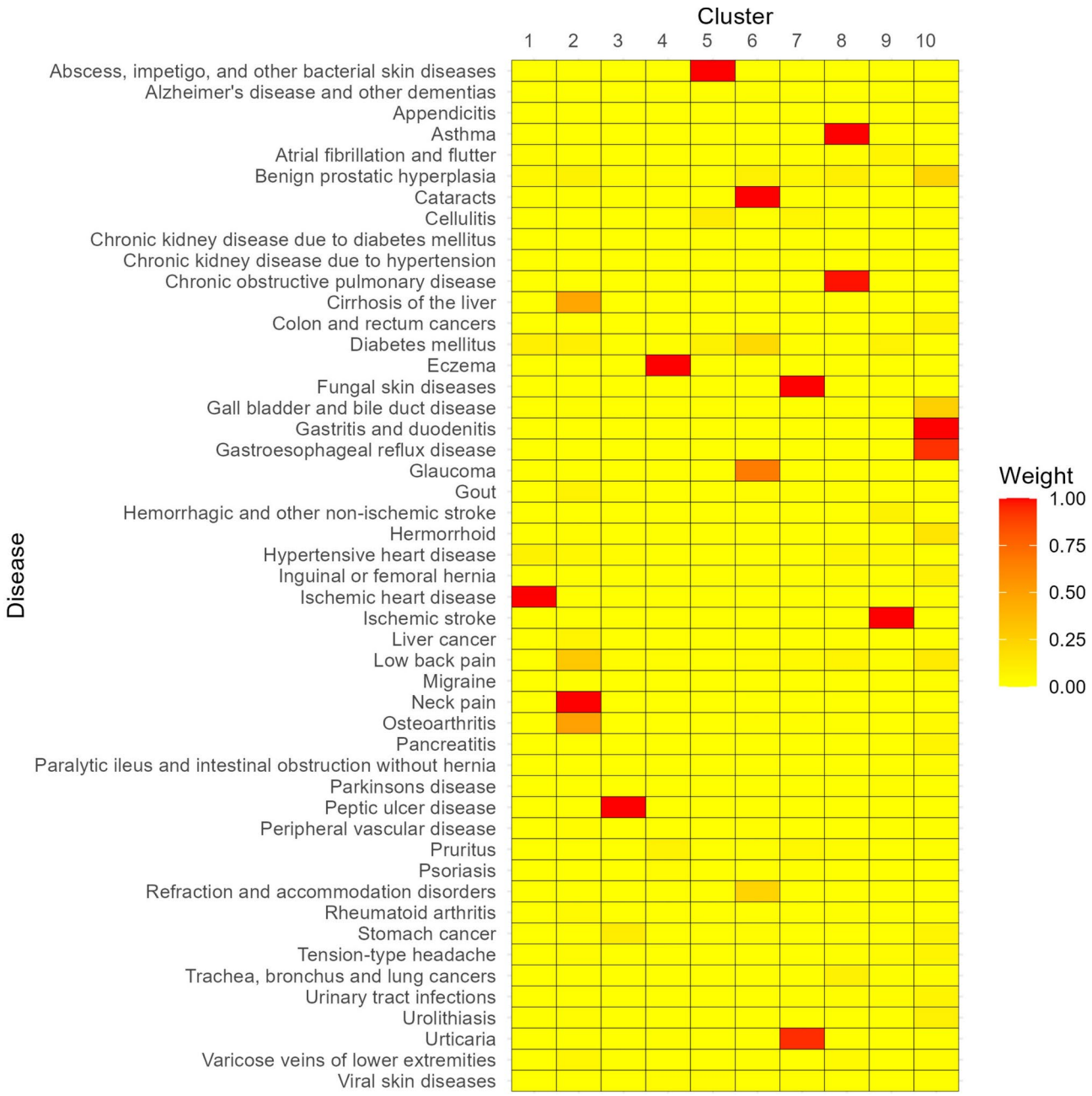

Fig. 1. Heatmap of NMF result for male in their 50 s.

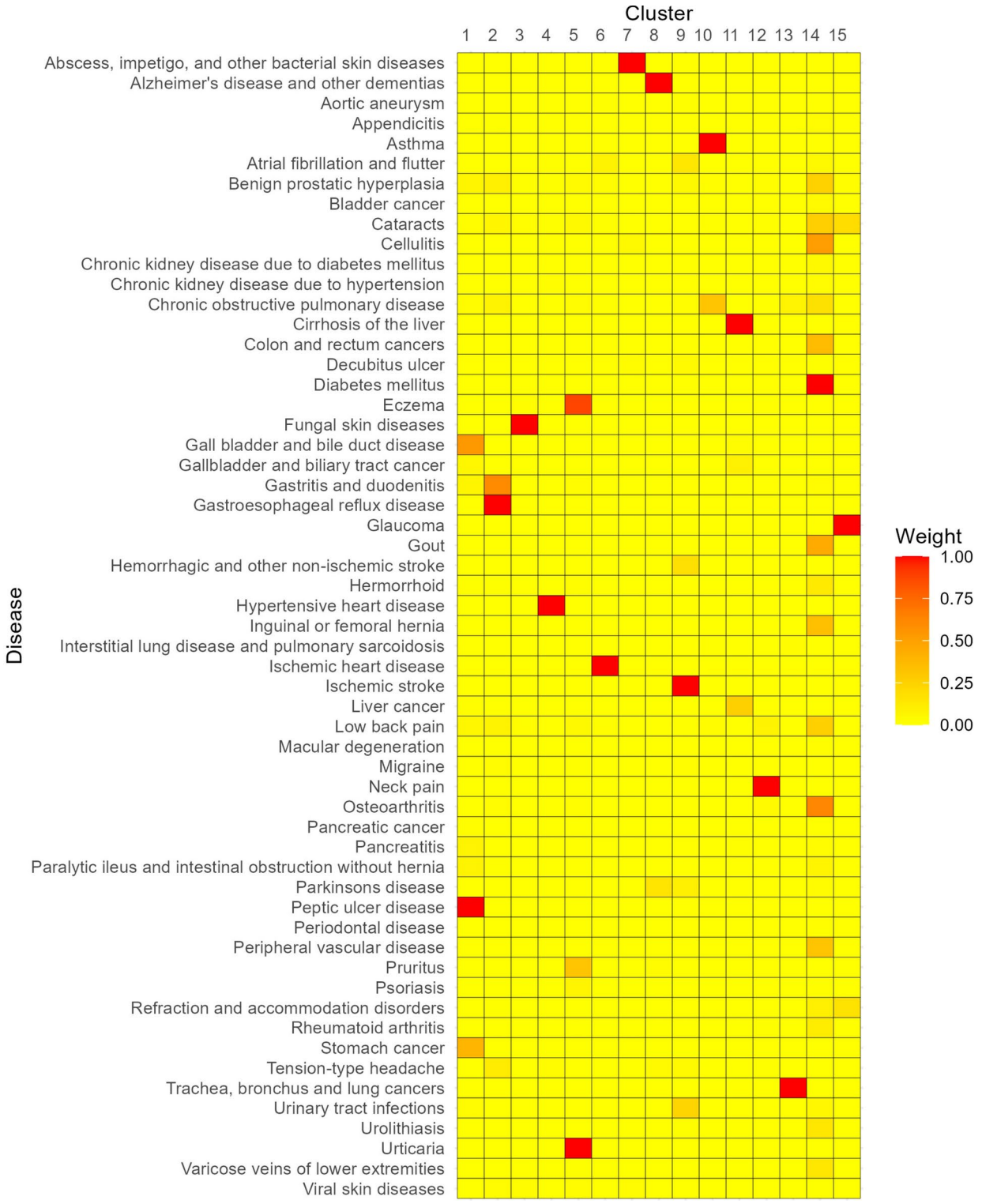

Fig. 2. Heatmap of NMF result for male in their 60 s.

- Cluster 4: Stroke, diabetes
- Clusters 5, 7, and 8: Skin and musculoskeletal diseases
- Cluster 6: COPD, musculoskeletal diseases
- Cluster 9: Urologic diseases
- Cluster 10: Gall bladder and bile duct disease, cirrhosis of the liver, pancreatitis
- Cluster 11: Skin diseases
- Cluster 12: Musculoskeletal diseases

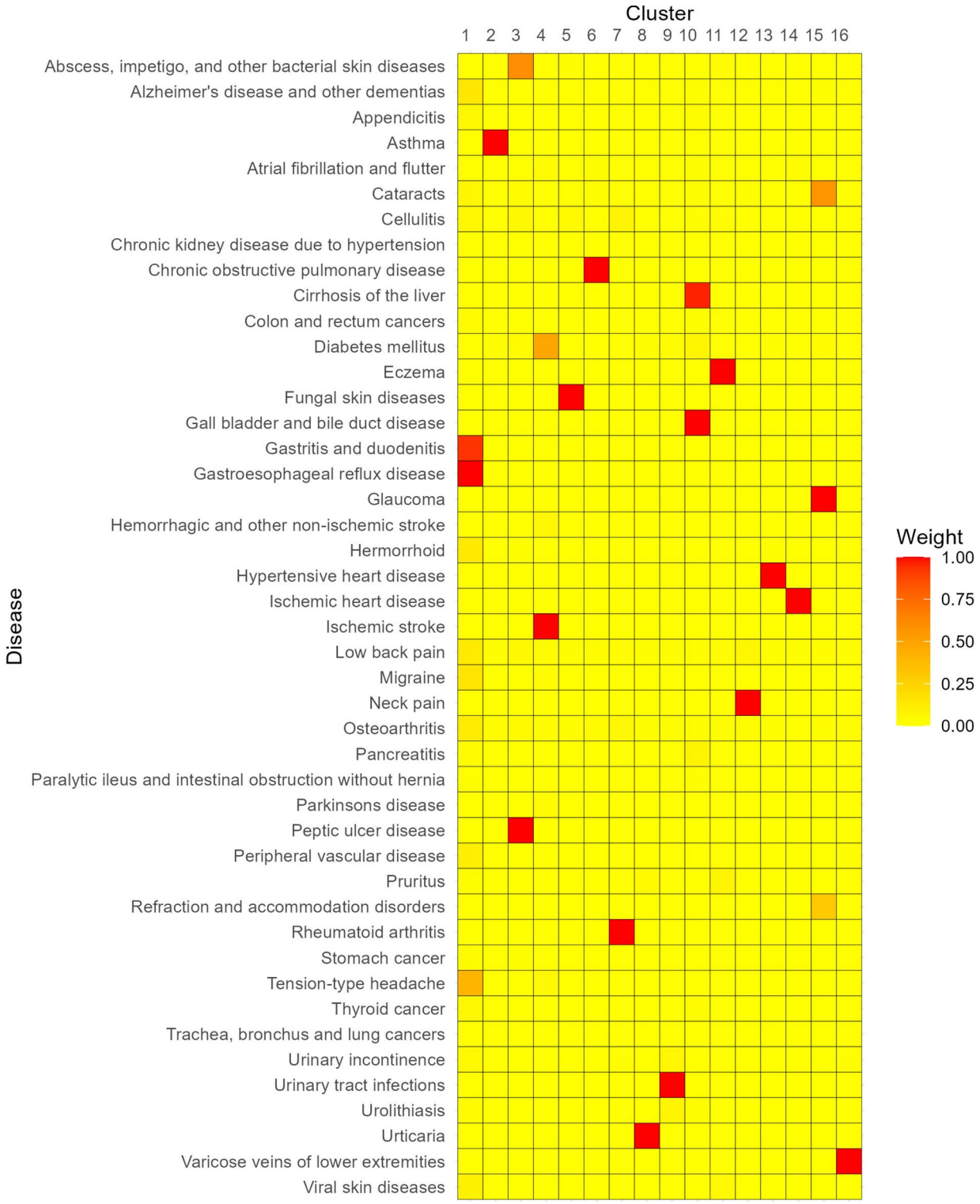

Fig. 3. Heatmap of NMF result for female in their 50 s.

- Cluster 13: Hypertensive heart disease, hemorrhoid, OA
- Cluster 14: IHD, AFF, LBP
- Cluster 15: Eye diseases
- Cluster 16: Varicose veins of lower extremities, OA, peripheral vascular disease (PVD)

EFA confirmed clusters 1, 11, and 12, while NMF identified additional novel clusters. Furthermore, OA appeared in 8 of the 16 identified clusters.

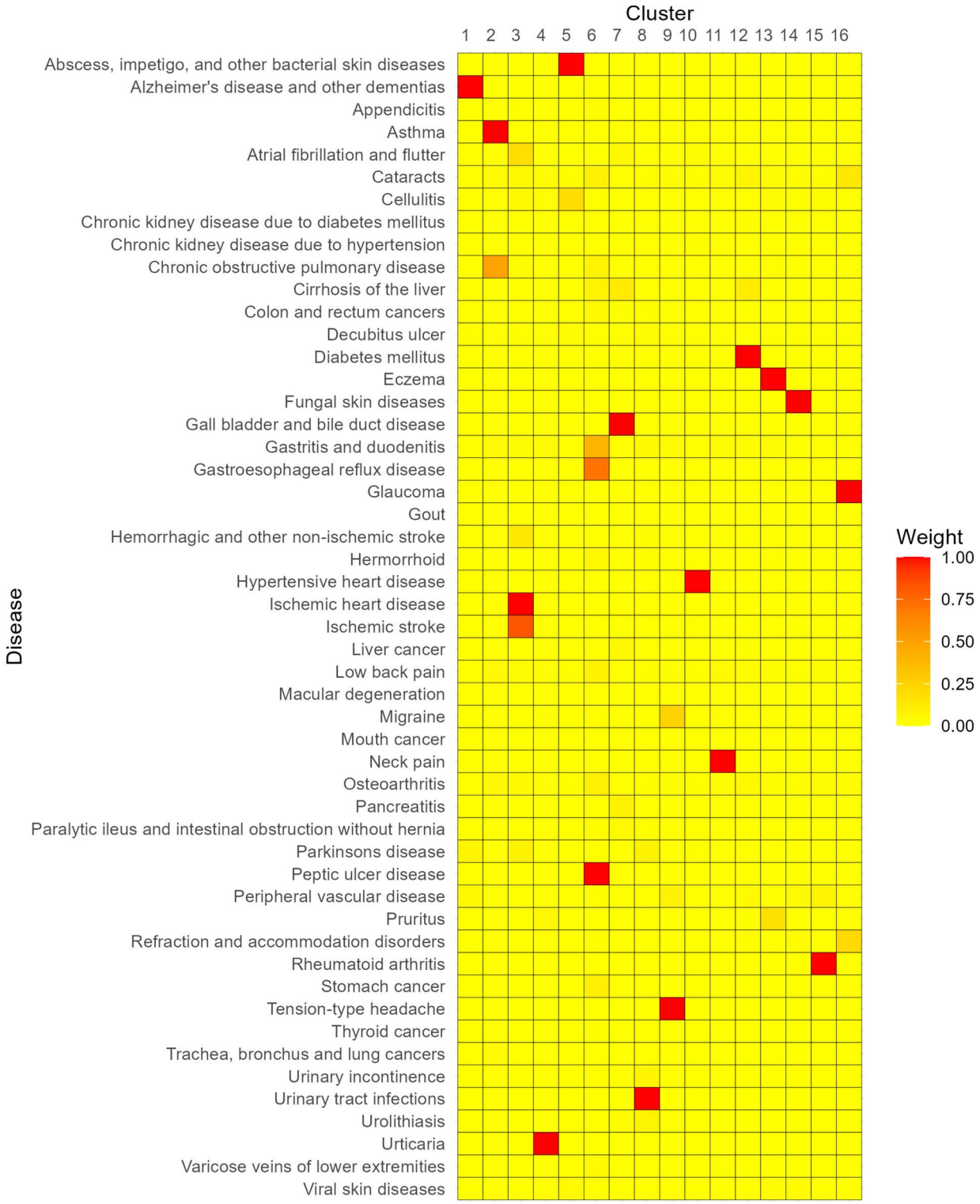

Fig. 4. Heatmap of NMF result for female in their 60 s.

Women in their 60 s The study identified 16 clusters among women in their 60 s (Table 3) (Fig. 4).

- Cluster 1: Alzheimer's disease, Parkinson's disease, PVD
- Cluster 2: Respiratory diseases, OA
- Cluster 3: IHD, stroke, AFF
- Cluster 4: Skin diseases, Gastrointestinal diseases
- Cluster 5 and 13: Skin diseases

- Cluster 6: Gastrointestinal diseases
- Cluster 7: Gall bladder and bile duct disease, cirrhosis of the liver, pancreatitis
- Cluster 8: Urologic diseases, Parkinson's disease
- Cluster 9: TTH, migraine, PVD
- Cluster 10: Hypertensive heart disease, AFF, Parkinson's disease
- Cluster 11: Musculoskeletal diseases, Gastrointestinal diseases
- Cluster 12: Diabetes, cirrhosis of the liver, cataracts
- Cluster 14: Skin and eye diseases
- Cluster 15: Musculoskeletal diseases, PVD
- Cluster 16: Eye diseases

EFA validated clusters 5, 6, and 13, while NMF revealed additional novel clusters. PVD was distributed across four clusters, and the combinations within cluster 7 mirrored those observed in cluster 10 among women in their 50 s.

After identifying patients and their respective time weights per cluster in matrix B, the disease onset age within each cluster was visualized using a box plot. Box plots visualized the age of disease onset for individuals within each cluster, displaying the mean, median, and interquartile range. Figure S2 displays the distribution of age of disease onset within clusters for male in their 50 s, highlighting the sequential development of PUD and LBP within cluster 3. Figure S3 illustrates the progression of hypertensive heart disease to CKD due to hypertension in a man in his 60 s (cluster 4). Furthermore, among men in their 60 s, cluster 6 showed IHD following AFF, cluster 8 revealed Parkinson's disease after Alzheimer's disease, cluster 9 indicated UTI following a stroke, and cluster 14 presented cellulitis after diabetes mellitus. Figure S4 illustrates the disease onset sequence in women in their 50 s, showing the progression from urolithiasis to UTI (cluster 9) and from AFF to IHD (cluster 14). Figure S5 illustrates disease progression among women in their 60 s, revealing that Alzheimer's disease also followed Parkinson's disease (cluster 1). Disease combinations in different age groups, such as AFF progressing to IHD (cluster 3) and urolithiasis preceding UTI (cluster 8), were also observed in women in their 60 s.

Discussion

In this study, we analyzed healthcare utilization data from 142,505 adults aged 50 to 69, spanning from 2002 to 2019, to examine the prevalence and incidence of 126 chronic diseases in Korea. In 2002, the total population of individuals aged 50 to 69 in South Korea was approximately 8 million, with our study covering 1.8% of this group. Our approach employed EFA to identify multimorbidity patterns in the prevalence of chronic diseases, alongside machine learning techniques like NMF to explore multimorbid patterns based on the sequence of disease incidence. NMF analysis revealed more disease clusters and a wider array of temporal disease combinations than EFA. The EFA analysis indicates that explained variance possesses limited explanatory power of 15 to 25% per subgroup, and this results from an analysis of actual data (Figures S6, S7). The same figure displays the explained variance ratio for NMF, which was determined by the number of clusters. NMF advantages such as enhanced interpretability that may offset the limitations of EFA.

The EFA analysis revealed that GERD had the highest factor loading in factor 1, while eczema had the highest loading in factor 2. Additionally, CKD due to diabetes exhibited the highest factor loading in factor 3 across all subgroups, except for women in their 50 s. In factor 5, ischemic stroke had the highest loading among individuals in their 50 s, whereas decubitus ulcer had the highest loading among those in their 60 s. The NMF results indicated that most clusters were dominated by a single disease with the highest weight, with a few notable exceptions. Specifically, cluster 10 for males in their 50 s, cluster 2 for males in their 60 s, and cluster 1 for males in their 50 s showed similar weights for gastritis and GERD. Additionally, cluster 4 for females in their 60 s was an exception, as both ischemic heart disease (IHD) and ischemic stroke exhibited high weights. However, the disease with the highest factor loading or weight within a cluster was not necessarily the most prevalent in that group, nor could it be conclusively identified as the index disease in comorbidity or as playing a mechanistically significant role in multimorbidity. This is because our study does not establish causal relationships or underlying mechanisms between diseases. Rather, it aims to clarify disease trajectories by identifying patterns of clinically significant diseases. Further investigation is warranted to explore the relationship between predominant diseases and their comorbid conditions to better understand their roles within multimorbidity clusters.

Identifying comorbidity patterns and understanding their underlying mechanisms are expected to drive the development of clinical and public health strategies for comorbidity prevention and management. These insights provide a foundation for advancing precision medicine and precision public health, both of which aim to prevent comorbidity. Traditional medical approaches, which treat each condition separately without considering comorbidities, fail to address the complexity of multimorbidity. This not only increases economic strain through duplicated care but also worsens outcomes due to conflicting treatments[45]. Identifying potentially associated disease pairs can alert clinicians to monitor for the subsequent occurrence of conditions in patients with existing diseases. This information serves as a crucial foundation for clinical practice aimed at preventing the development of comorbidities, thereby reducing the burden associated with multiple chronic conditions. Furthermore, additional research on the comorbidity patterns identified in this study could help determine whether direct causal relationships exist between diseases, elucidate underlying mechanisms, or identify whether common risk factors are shared between the two diseases, significantly advancing precision preventive medicine. Thus, healthcare and public health policies should integrate multimorbidity considerations rather than focusing solely on individual diseases. Since this study analyzed healthcare utilization data from South Koreans, the observed patterns may be specific to Korean cultural and environmental characteristics. Investigating comorbidity patterns in diverse healthcare systems and socioeconomic contexts could enhance the generalizability of these findings.

Our findings suggest that certain diseases co-occur, with distinct comorbidity patterns observed across sexes and age groups. These results highlight the importance of considering age- and sex-specific patterns in understanding comorbidity clusters, potentially informing tailored interventions and management strategies. This investigation aimed to discern chronic disease patterns within individual patients over a prolonged follow-up period, revealing both the co-occurrence of multiple conditions and the sequence in which they may appear. The identification of established disease pairs adds to the study's validity, while the discovery of less familiar comorbidity patterns opens new avenues for research.

The study uncovered previously unknown or under-researched multimorbidity patterns between diseases, shedding light on potential underlying mechanisms contributing to their co-occurrence. Identifying these novel comorbidity patterns highlights the necessity for further research to investigate potential causal relationships between these diseases. It also suggests which disease might precede the other in a bidirectional relationship. These findings also offer crucial clinical insights, informing preventative strategies and mitigating the burden of comorbidities, thereby emphasizing the study's contribution to understanding multimorbidity and informing clinical strategies.

GERD, gastritis and duodenitis, and PUD were consistently comorbid with TTH across all groups, with an additional comorbidity with migraine observed in men in the EFA group. This comorbidity may be explained by common pathophysiological mechanisms involving the gut-brain axis. Disruptions in the glutamate pathway within the enteric nervous system and the gut-brain axis are linked to gastrointestinal disorders characterized by hyperacidity, such as GERD, gastritis, duodenal ulcer, and gastric ulcer. Concurrently, glutamate plays a crucial role in migraine pathogenesis, with elevated levels observed in the plasma and cerebrospinal fluid of migraine patients. Furthermore, among the extraintestinal pathologies associated with *H. pylori* infection, a well-known risk factor for gastritis and PUD, reports have linked it to migraine and TTH[46–50].

Another comorbid pair identified is COPD and GERD. COPD-induced respiratory changes, including hyperinflation and increased inspiratory effort, impact the lower esophageal sphincter, promoting reflux[51]. Simultaneously, GERD worsens COPD symptoms, influenced by factors such as smoking, medications, cardiac complications, sleep apnea, elevated body mass index, and mechanical changes from COPD[51–55]. Our findings highlight the need for public health policies addressing shared risk factors, emphasizing the importance of targeted strategies informed by comorbidity clusters.

Additionally, a comorbidity between Parkinson's disease and ischemic stroke was identified, though its precise nature is unclear. One hypothesized mechanism involves oxidative stress in Parkinson's disease, which disrupts cellular homeostasis[56, 57]. This stress may damage dopaminergic cells and induce endothelial atherosclerotic changes, raising stroke risk. The high oxidative stress levels associated with Parkinson's disease may contribute to a higher stroke incidence. Furthermore, Parkinson's disease-related autonomic dysfunction, leading to orthostatic hypotension, correlates with ischemic brain injury[58].

Our analysis revealed an association between Parkinson's disease and subsequent Alzheimer's disease development in both sexes during their 60 s within the NMF cohort. This ordering is evident in clusters 1 and 9, as shown in Figures S3 and S5. Cognitive impairment, often culminating in dementia, is a frequent complication of Parkinson's disease[59]. While Parkinson's is associated with alpha-synucleinopathy and Alzheimer's with beta-amyloid and tauopathy, both share abnormal neuronal accumulation of toxic substances, potentially leading to overlapping pathologies[60]. Genetic links between Alzheimer's and Parkinson's diseases have also been reported[61]. Additionally, Alzheimer's dementia codes may have been applied to other forms of dementia associated with Parkinson's disease.

Lastly, the EFA and NMF groups exhibited comorbidities of ischemic stroke, hemorrhagic stroke, and other non-ischemic strokes across all age groups, except for women in their 60 s in the NMF group. While no direct causal relationship exists between these conditions, common risk factors such as hypertension, diabetes, obesity, physical inactivity, dyslipidemia, cardiovascular disease, smoking, and alcohol consumption may explain this comorbidity[62].

This study reaffirmed established associations between specific conditions. The comorbidity of cirrhosis of the liver and liver cancer was observed in men in their 50 s and 60 s in EFA and in men in their 60 s in NMF. Cirrhosis, characterized by chronic inflammation and scarring due to ongoing liver injury, creates an environment conducive to liver cancer development by promoting genetic abnormalities and cellular dysfunction[63].

The consistent comorbidity of GERD, gastritis, and duodenitis across all sexes and ages can be attributed to a shared mechanism of hyperacidity. These conditions are influenced by similar risk factors, including alcohol use, smoking, NSAIDs, *Helicobacter pylori* infection, and certain dietary habits[64, 65]. This suggests that targeted interventions addressing these factors could reduce the prevalence of these conditions.

The comorbid pattern of ischemic stroke and AFF was observed in males in their 50 s in the EFA and in females in their 60 s in the NMF. The risk of ischemic stroke is significantly increased by AFF and vice versa. AFF causes blood stasis in the atria, leading to thrombus formation, particularly in the left atrial appendage[66, 67]. These thrombi can dislodge and occlude cerebral arteries, resulting in ischemic strokes[67, 68].

This study also observed the comorbidity of AFF and IHD in women aged 60 and above in the EFA group, men aged 60 and above, and women of all ages in the NMF group. AFF and IHD are recognized to have a bidirectional relationship. However, this study revealed that AFF was followed by IHD, as seen by cluster 6 in Figure S3, cluster 14 in Figure S4, and cluster 3 in Figure S5. One possible rationale is that AFF may contribute to IHD by increasing heart rate and oxygen consumption, thereby accelerating atherosclerosis[69].

The comorbid pattern of IHD and hypertensive heart disease was observed in women in their 60 s in the EFA and men in their 50 s in the NMF. Chronic high blood pressure leads to left ventricular hypertrophy and increases myocardial oxygen demand, exacerbating IHD[70].

Men in their 60 s in the NMF and women in their 60 s in the EFA exhibited a comorbid pattern of hypertensive heart disease and CKD due to hypertension. These conditions have a bidirectional relationship, highlighting the

interconnected nature of chronic diseases. Hypertensive heart disease can lead to CKD by restricting blood flow to the kidneys and promoting arteriosclerosis. Conversely, CKD exacerbates hypertensive heart disease through mechanisms like hypertension, fluid retention, and metabolic disruptions[71].

A comorbid pattern of gout and CKD was also identified among women in their 60 s in the EFA. Hyperuricemia contributes to kidney injury through mechanisms including endothelial dysfunction, oxidative stress, and inflammation[72, 73].

The coexistence of asthma and COPD was observed across various age groups in both the EFA and NMF. Asthma-COPD overlap involves variable airway obstruction from asthma and persistent airflow limitation from COPD due to smoking-induced inflammation[74]. These conditions also share genetic associations and allergic inflammation[75].

An association between OA and GERD was found in male groups within the EFA. OA is characterized by chronic low-grade inflammation, which can contribute to GERD[76]. Additionally, the use of NSAIDs for OA treatment damages the gastric mucosa and increases acid secretion, heightening GERD risk. Similarly, comorbidity of LBP and digestive system diseases was noted in multiple groups.

In the 60 s age group, a comorbidity of Parkinson's disease, stroke, and decubitus ulcers was observed in the EFA. Both Parkinson's disease and ischemic stroke are associated with an increased risk of decubitus ulcers due to the higher likelihood of being bedridden[77].

Diabetes mellitus showed comorbidity with multiple diseases, including IHD, hypertensive heart disease, stroke, OA, asthma, skin diseases, liver diseases, and eye diseases, across all age and sex groups. Diabetes mellitus leads to elevated blood sugar levels, which damage blood vessels and nerves[78, 79]. Additionally, chronic inflammation intensifies insulin resistance, worsening diabetes[80, 81]. This increases the risk of heart disease, infections, strokes, and joint problems.

Certain diseases appear in multiple clusters due to their high prevalence. For example, BPH, most prevalent among men, emerged in multiple clusters derived from the NMF of men in their 50 s. Similarly, OA was present in multiple clusters for women in their 50 s. Other musculoskeletal conditions, such as LBP and neck pain, also appeared in these clusters, likely reflecting their high prevalence despite a lack of strong clinical associations with other conditions in the same clusters.

Our study presents several notable strengths compared to previous research on multimorbidity patterns. First, we included a substantial sample size of 142,505 patients and analyzed 126 chronic conditions to capture diverse interpretations. By leveraging big data, we incorporated comorbid cases and identified even minor comorbid patterns. Many prior studies have investigated comorbidity patterns through various analyses but were limited by the number of subjects and diseases considered, thereby minimizing the detection of disease patterns. Additionally, we utilized longitudinal data, tracking each individual for nearly 20 years. Given that many disease co-occurrences and interactions take years to develop, the longitudinal nature of our data is crucial for drawing meaningful conclusions. Furthermore, we aimed to explore not only patterns of multimorbidity but also the chronological sequence of comorbid diseases by considering temporal aspects, employing both traditional and machine learning methods. Most previous research used pairwise methods, which assess diseases in pairs, or factorization methods, which extract latent factors. However, they did not account for the temporal evolution of diseases over time. We sought to examine the temporal characteristics of disease clusters to generate new hypotheses about potential causal pathways between diseases. By elucidating these temporal comorbidity patterns, we aim to provide valuable insights into disease prevention, which can significantly enhance clinical practice and reduce the burden of comorbidity.

However, this study has several limitations that require careful consideration. First, the study utilized claims data, which may not fully represent a patient's true morbidity due to potential inaccuracies in ICD-10 coding. Additionally, there may be a time discrepancy between when a patient seeks medical attention for a disease and when the disease occurs. As a result, several comorbid pairs showed a sequence of comorbidity that differs from the commonly known order of disease progression. For example, hypertensive heart disease was observed to occur after AFF in women in their 60 s, and diabetes mellitus was noted to develop after IHD in men in their 50 s. Additionally, diabetes mellitus was found to develop following a stroke in individuals of both sexes in their 50 s. Therefore, future research should incorporate additional information from tests, drugs, and treatments to improve the accuracy of disease prevalence and incidence estimates. Furthermore, healthcare utilization may be biased, potentially omitting mild cases due to the patients' infrequent hospital visits. Consequently, the interpretation of factors and clusters should be approached with caution. Second, the identified comorbidity patterns do not necessarily indicate causal relationships between the grouped diseases and should be interpreted with caution. Finally, both hypertension and dyslipidemia were excluded from this analysis. Excluding these conditions may have affected the observed results. However, this study included hypertensive heart disease was listed among the diseases rather than hypertension, which may indirectly elucidate the impact of hypertension on the observed patterns. Excluding established risk factors allowed us to focus on potentially novel or less reviewed combinations of diseases.

Nevertheless, identifying potentially associated disease pairs can alert clinicians to monitor for the subsequent occurrence of conditions in patients with existing diseases. This information can serve as a crucial basis for clinical practice aimed at preventing the development of comorbidities, thereby reducing the burden associated with multiple chronic conditions. Moreover, further epidemiological and clinical research on the comorbidity patterns identified in this study could help determine whether causal relationships exist between diseases and elucidate underlying mechanisms, significantly benefiting precision preventive medicine.

Conclusion

This study identifies distinct multimorbidity patterns, providing actionable insights into the co-occurrence of chronic diseases. The accumulation of such evidence has the potential to guide targeted interventions and

inform healthcare policies, enabling healthcare providers to predict disease trajectories and intervene proactively in the development of comorbidities. A personalized approach to managing multimorbidity offers a strategic opportunity to improve patient outcomes and quality of life, providing a strategic response to the growing prevalence of chronic multimorbidity.

Data availability

The data supporting the findings of this study were provided by the National Health Insurance Service (NHIS) and were used under license for this study, with the restriction that they are not publicly available. However, the data are available from the authors upon reasonable request and with permission from the NHIS. For any inquiries, please contact the corresponding author.

Received: 30 September 2024; Accepted: 13 March 2025
Published online: 22 March 2025

References

1. Raghupathi, W. & Raghupathi, V. An Empirical Study of Chronic Diseases in the United States: A Visual Analytics Approach. Int J Environ Res Public Health 15 (2018).
2. Anderson, E. & Durstine, J. L. Physical activity, exercise, and chronic diseases: A brief review. *Sports Med. Health Sci.* **1**, 3–10 (2019).
3. OECD. Health at a Glance 2023. OECD Indicators (2023).
4. Park, B. et al. Multimorbidity and health-related quality of life in Koreans aged 50 or older using KNHANES 2013–2014. *Health Qual. Life Outcomes* **16**, 186 (2018).
5. Garin, N. et al. Impact of multimorbidity on disability and quality of life in the Spanish older population. *PLoS One* **9**, e111498 (2014).
6. Barnett, K. et al. Epidemiology of multimorbidity and implications for health care, research, and medical education: a cross-sectional study. *Lancet* **380**, 37–43 (2012).
7. Johnston, M. C., Crilly, M., Black, C., Prescott, G. J. & Mercer, S. W. Defining and measuring multimorbidity: a systematic review of systematic reviews. *Eur. J. Public Health* **29**, 182–189 (2019).
8. Sav, A. et al. Burden of treatment for chronic illness: a concept analysis and review of the literature. *Health Expect.* **18**, 312–324 (2015).
9. Marengoni, A. et al. Aging with multimorbidity: a systematic review of the literature. *Ageing Res. Rev.* **10**, 430–439 (2011).
10. Álvarez-Gálvez, J. et al. Social determinants of multimorbidity patterns: A systematic review. *Front. Public Health* **11**, 1081518 (2023).
11. Ju, H.-H. Improving care coordination of patients with chronic diseases. *J. Nurse Pract.* **18**, 833–836 (2022).
12. Peer, N., de Villiers, A., Jonathan, D., Kalombo, C. & Kengne, A. P. Care and management of a double burden of chronic diseases: Experiences of patients and perceptions of their healthcare providers. *PLoS One* **15**, e0235710 (2020).
13. Chen, H.-L. et al. Multimorbidity patterns and the association with health status of the oldest-old in long-term care facilities in China: A two-step analysis. *BMC Geriatr.* **23**, 851 (2023).
14. Fan, J., Sun, Z., Yu, C., Guo, Y., Pei, P., Yang, L. et al. Multimorbidity patterns and association with mortality in 0.5 million Chinese adults. Chin. Med. J. (Engl) 135, 648–657 (2022).
15. Forslund, T., Carlsson, A. C., Ljunggren, G., Ärnlöv, J. & Wachtler, C. Patterns of multimorbidity and pharmacotherapy: A total population cross-sectional study. *Fam. Pract.* **38**, 132–140 (2021).
16. Guisado-Clavero, M. et al. Multimorbidity patterns in the elderly: a prospective cohort study with cluster analysis. *BMC Geriatr.* **18**, 16 (2018).
17. Lleal, M. et al. Multimorbidity patterns in COVID-19 patients and their relationship with infection severity: MRisk-COVID study. *PLoS One* **18**, e0290969 (2023).
18. Tazzeo, C. et al. Multimorbidity patterns and risk of frailty in older community-dwelling adults: a population-based cohort study. *Age Ageing* **50**, 2183–2191 (2021).
19. Violán, C. et al. Multimorbidity patterns with K-means nonhierarchical cluster analysis. *BMC Fam Pract.* **19**, 108 (2018).
20. Carretero-Bravo, J., Ramos-Fiol, B., Ortega-Martín, E., Suárez-Lledó, V., Salazar, A., O'Ferrall-González, C. et al. Multimorbidity patterns and their association with social determinants, mental and physical health during the COVID-19 pandemic. Int. J. Environ. Res. Public Health 19, (2022).
21. Liu, Y. & Jiang, D. Multimorbidity patterns in US adults with subjective cognitive decline and their relationship with functional difficulties. *J. Aging Health* **34**, 929–938 (2022).
22. Lu, J. et al. Multimorbidity patterns in old adults and their associated multi-layered factors: a cross-sectional study. *BMC Geriatr.* **21**, 372 (2021).
23. Roomaney, R. A., van Wyk, B., Cois, A. & Pillay van-Wyk, V. Multimorbidity patterns in South Africa: A latent class analysis. Front Public Health 10, 1082587 (2022).
24. Zhong, Y. et al. Prevalence, patterns of multimorbidity and associations with health care utilization among middle-aged and older people in China. *BMC Public Health* **23**, 537 (2023).
25. Aoki, T., Yamamoto, Y., Ikenoue, T., Onishi, Y. & Fukuhara, S. Multimorbidity patterns in relation to polypharmacy and dosage frequency: a nationwide, cross-sectional study in a Japanese population. *Sci. Rep.* **8**, 3806 (2018).
26. England, B. R. et al. Identification of multimorbidity patterns in rheumatoid arthritis through machine learning. *Arthritis Care Res. (Hoboken)* **75**, 220–230 (2023).
27. Honda, Y., Nakamura, M., Aoki, T. & Ojima, T. Multimorbidity patterns and the relation to self-rated health among older Japanese people: a nationwide cross-sectional study. *BMJ Open* **12**, e063729 (2022).
28. Wang, Z. et al. Association between multimorbidity patterns and disability among older people covered by long-term care insurance in Shanghai. *China. BMC Public Health* **21**, 418 (2021).
29. Damiano, C. et al. Frailty, multimorbidity patterns and mortality in institutionalized older adults in Italy. *Aging Clin Exp Res* **34**, 3123–3130 (2022).
30. Roso-Llorach, A. et al. 12-year evolution of multimorbidity patterns among older adults based on Hidden Markov Models. *Aging (Albany NY)* **14**, 9805–9817 (2022).
31. Chen, Y., Shi, L., Zheng, X., Yang, J., Xue, Y. & Xiao, S. et al. Patterns and determinants of multimorbidity in older adults: study in health-ecological perspective. Int J Environ Res Public Health 19, (2022).
32. 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 17, (2020).
33. Rodrigues, L. P. et al. Multimorbidity patterns and hospitalisation occurrence in adults and older adults aged 50 years or over. *Sci. Rep.* **12**, 11643 (2022).

34. Lee, J., Lee, J. S., Park, S. H., Shin, S. A. & Kim, K. Cohort Profile: The National Health Insurance Service-National Sample Cohort (NHIS-NSC). *South Korea. Int. J. Epidemiol.* **46**, e15 (2017).
35. Yoon, S. A Study on the Development of a Health Status Measurement Platform for Implementation of Precision Public Health in Korea. Ministry of Health and Welfare (2022).
36. Correia, E. T. O. et al. The hypertension-based chronic disease model in a primary care setting. *Int J Cardiol Cardiovasc Risk Prev.* **18**, 200204. https://doi.org/10.1016/j.ijcrp.2023.200204.PMID:37664167;PMCID:PMC10470300 (2023).
37. Addisu, B., Bekele, S., Wube, T. B., Hirigo, A. T. & Cheneke, W. Dyslipidemia and its associated factors among adult cardiac patients at Ambo university referral hospital, Oromia region, west Ethiopia. *BMC cardiovascular disorders* **23**(1), 321. https://doi.org/10.1186/s12872-023-03348-y (2023).
38. Mikolasevic, I., Žutelija, M., Mavrinac, V. & Orlic, L. Dyslipidemia in patients with chronic kidney disease: etiology and management. *International journal of nephrology and renovascular disease* **10**, 35–45. https://doi.org/10.2147/IJNRD.S101808 (2017).
39. Yao, S. S. et al. Associations between multimorbidity and physical performance in older Chinese adults. *Int. J. Environ. Res. Public Health* **17**, 4546 (2020).
40. Prados-Torres, A. et al. Multimorbidity patterns in primary care: interactions among chronic diseases using factor analysis. *PLoS One* **7**, e32190 (2012).
41. Xia, M., Li, S., Chen, W. & Yang, G. Chapter Five - Perceptual image hashing using rotation invariant uniform local binary patterns and color feature. In Advances in Computers (ed. Hurson, A. R.) 130, 163–205 (Elsevier, 2023).
42. Hassaine, A. et al. Learning multimorbidity patterns from electronic health records using Non-negative Matrix Factorisation. *J. Biomed. Inform.* **112**, 103606 (2020).
43. Lee, S. L. et al. Computed tomography and ultrasonography do not improve and may delay the diagnosis and treatment of acute appendicitis. *Arch. Surg.* **136**, 556–562 (2001).
44. Lee, D. D. & Seung, H. S. Learning the parts of objects by non-negative matrix factorization. *Nature* **401**, 788–791 (1999).
45. Skou, S. T. et al. *Multimorbidity. Nat. Rev. Dis. Primers* **8**, 48 (2022).
46. Milovanovic, B. et al. Cardiac autonomic dysfunction in patients with gastroesophageal reflux disease. *World J. Gastroenterol.* **21**, 6982–6989 (2015).
47. Arzani, M. et al. Gut-brain axis and migraine headache: a comprehensive review. *J. Headache Pain* **21**, 15 (2020).
48. Ruszniewski, P., Soufflet, C. & Barthelemy, P. Nonsteroidal anti-inflammatory drug use as a risk factor for gastro-oesophageal reflux disease: an observational study. *Aliment. Pharmacol. Ther.* **28**, 1134–1139 (2008).
49. Lee, S. H., Lee, J. J., Kwon, Y., Kim, J. H. & Sohn, J. H. Clinical implications of associations between headache and gastrointestinal disorders: A study using the Hallym Smart Clinical Data Warehouse. *Front. Neurol.* **8**, 526 (2017).
50. M, T. N., Rezaeizadeh, H., Fazljoo, S. M. & Keshavarz, M. Gastrointestinal headache; a narrative review. Emerg (Tehran) 4, 171–183 (2016).
51. Pauwels, A., Blondeau, K., Dupont, L. J. & Sifrim, D. Mechanisms of increased gastroesophageal reflux in patients with cystic fibrosis. *Am. J. Gastroenterol.* **107**, 1346–1353 (2012).
52. Gadel, A. A., Mostafa, M., Younis, A. & Haleem, M. Esophageal motility pattern and gastro-esophageal reflux in chronic obstructive pulmonary disease. *Hepatogastroenterology* **59**, 2498–2502 (2012).
53. Martinez, C. H. et al. Impact of self-reported gastroesophageal reflux disease in subjects from COPDGene cohort. *Respir. Res.* **15**, 62 (2014).
54. Pandolfino, J. E. & Kahrilas, P. J. Smoking and gastro-oesophageal reflux disease. *Eur. J. Gastroenterol. Hepatol.* **12**, 837–842 (2000).
55. Ruzkowski, C. J., Sanowski, R. A., Austin, J., Rohwedder, J. J. & Waring, J. P. The effects of inhaled albuterol and oral theophylline on gastroesophageal reflux in patients with gastroesophageal reflux disease and obstructive lung disease. *Arch. Intern. Med.* **152**, 783–785 (1992).
56. Kim, W. Y. et al. Association of Parkinson's disease with ischemic stroke in Korea: A nationwide longitudinal cohort study in Korea. *J. Cerebrovasc. Endovasc. Neurosurg.* **23**, 233–239 (2021).
57. Subramaniam, S. R. & Chesselet, M. F. Mitochondrial dysfunction and oxidative stress in Parkinson's disease. *Prog. Neurobiol.* **106–107**, 17–32 (2013).
58. McDonald, C., Newton, J. L. & Burn, D. J. Orthostatic hypotension and cognitive impairment in Parkinson's disease: Causation or association?. *Mov. Disord.* **31**, 937–946 (2016).
59. Aarsland, D. & Kurz, M. W. The epidemiology of dementia associated with Parkinson disease. *J. Neurol. Sci.* **289**, 18–22 (2010).
60. Clinton, L. K., Blurton-Jones, M., Myczek, K., Trojanowski, J. Q. & LaFerla, F. M. Synergistic interactions between Abeta, tau, and alpha-synuclein: acceleration of neuropathology and cognitive decline. *J. Neurosci.* **30**, 7281–7289 (2010).
61. Liu, G. et al. Identifying the association between Alzheimer's disease and Parkinson's disease using genome-wide association studies and protein-protein interaction network. *Mol. Neurobiol.* **52**, 1629–1636 (2015).
62. Alkahtani, R. Molecular mechanisms underlying some major common risk factors of stroke. *Heliyon* **8**, e10218 (2022).
63. Kanda, T., Goto, T., Hirotsu, Y., Moriyama, M. & Omata, M. Molecular mechanisms driving progression of liver cirrhosis towards hepatocellular carcinoma in chronic hepatitis B and C infections: A review. Int. J. Mol. Sci. 20, (2019).
64. Finocchio, E. et al. Gastritis and gastroesophageal reflux disease are strongly associated with non-allergic nasal disorders. *BMC Pulm. Med.* **21**, 53 (2021).
65. Sugano, K. et al. Kyoto global consensus report on Helicobacter pylori gastritis. *Gut* **64**, 1353–1367 (2015).
66. Cresti, A. et al. Prevalence of extra-appendage thrombosis in non-valvular atrial fibrillation and atrial flutter in patients undergoing cardioversion: a large transoesophageal echo study. *EuroIntervention* **15**, e225–e230 (2019).
67. Migdady, I., Russman, A. & Buletko, A. B. Atrial fibrillation and ischemic stroke: A clinical review. *Semin. Neurol.* **41**, 348–364 (2021).
68. Hou, L. et al. Exploring the causal pathway from ischemic stroke to atrial fibrillation: a network Mendelian randomization study. *Mol. Med.* **26**, 7 (2020).
69. Oancea A. F., Jigoranu R. A., Morariu P. C., Miftode R. S., Trandabat B. A., Iov D. E., Cojocaru E., Costache I. I., Baroi L. G., Timofte D. V., Tanase D. M., & Floria M. Atrial Fibrillation and Chronic Coronary Ischemia: A Challenging Vicious Circle. Life (Basel). 13, 1370 (2023).
70. Masenga, S. K. & Kirabo, A. Hypertensive heart disease: risk factors, complications and mechanisms. *Front. Cardiovasc. Med.* **10**, 1205475 (2023).
71. Burnier, M. & Damianaki, A. Hypertension as Cardiovascular Risk Factor in Chronic Kidney Disease. *Circ. Res.* **132**, 1050–1063 (2023).
72. Singh, J. A. & Cleveland, J. D. Gout is associated with a higher risk of chronic renal disease in older adults: a retrospective cohort study of U.S. Medicare population. BMC Nephrol. 20, 93 (2019).
73. Busso, N. & So, A. Mechanisms of inflammation in gout. *Arthritis Res. Ther.* **12**, 206 (2010).
74. Dey, S., Eapen, M. S., Chia, C., Gaikwad, A. V., Wark, P. A. B. & Sohal, S. S. Pathogenesis, clinical features of asthma COPD overlap, and therapeutic modalities. Am. J. Physiol. Lung Cell Mol. Physiol. 322, L64-L83 (2022).
75. Itabashi, S. et al. Allergic sensitization in elderly patients with chronic obstructive pulmonary disease. *Respiration* **57**, 384–388 (1990).
76. Patel, J. et al. Predictors of Co-occurring Cardiovascular and Gastrointestinal Disorders among Elderly with Osteoarthritis. *Osteoarthr. Cartil. Open* **3**, 100148 (2021).

77. Jaul, E., Barron, J., Rosenzweig, J. P. & Menczel, J. An overview of co-morbidities and the development of pressure ulcers among older adults. *BMC Geriatr.* **18**, 305 (2018).
78. Lu, Y., Wang, W., Liu, J., Xie, M., Liu, Q. & Li, S. Vascular complications of diabetes: A narrative review. Medicine (Baltimore) 102, e35285 (2023).
79. Tomic, D., Shaw, J. E. & Magliano, D. J. The burden and risks of emerging complications of diabetes mellitus. *Nat. Rev. Endocrinol.* **18**, 525–539 (2022).
80. Lontchi-Yimagou, E., Sobngwi, E., Matsha, T. E. & Kengne, A. P. Diabetes mellitus and inflammation. *Curr. Diab. Rep.* **13**, 435–444 (2013).
81. Tsalamandris, S. et al. The role of inflammation in diabetes: Current concepts and future perspectives. *Eur. Cardiol.* **14**, 50–59 (2019).

Author contributions
Y.K. took part in both the data acquisition and analysis and drafted the work. S.P., Y.C., B.Y., S.K., J.P., and H.O. participated in the interpretation of the data. Y.L., and J.L. was involved in the design of the work. B.P. made substantial contribution to the conception, design of the work, and substantively revised the work. All authors read and approved the final manuscript.

Funding
This work was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (NRF- 2021R1F1A1063135).

Declarations

Competing interest
The authors declare no competing interests.

Ethical approval
The Institutional Review Board of Chung-Ang University waived the requirement for ethical approval for this study, which is a retrospective study and used anonymized data in accordance with the Bioethics and Safety Act (1041078–202112-HR-336–01). The authors assert that all procedures contributing to this work comply with the ethical standards of the relevant national and institutional committees on human experimentation and with the Helsinki Declaration of 1975, as revised in 2000.

Additional information
Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-025-94338-x.

Correspondence and requests for materials should be addressed to B.P.