



The Use of a Network Analysis to Identify Associations and Temporal Patterns Among Non-communicable Diseases in Japan Based on a Large Medical Claims Database

Shingo Higa¹ · Kazutaka Nozawa¹ · Yusuke Karasawa¹ · Chikako Shirai¹ · Satoshi Matsuyama¹ · Yuji Yamamoto² · Thomas Laurent³ · Yuko Asami¹

Accepted: 15 May 2022 / Published online: 2 July 2022
© The Author(s) 2022

Abstract

Background Reducing the considerable non-communicable disease (NCD) burden in the aging Japanese population depends on better understanding of the comorbid and temporal relationships between different NCDs.

Objective We aimed to identify associations between NCDs and temporal patterns of NCDs in Japan using data from a large medical claims database.

Methods The study used three-digit International Classification of Diseases, Tenth Revision codes for NCDs for employees and their dependents included in the MinaCare database, which covers the period since 2010. Associations between pairs of NCDs were assessed by calculating risk ratios. The calculated risk ratios were used to create a network of closely associated NCDs (risk ratio > 15, statistically significant) and to assess temporal patterns of NCD diagnoses (risk ratio \geq 5). The Infomap algorithm was used to identify clusters of diseases for different sex and age strata.

Results The analysis included 4,200,254 individuals (age < 65 years: 98%). Many of the temporal associations and patterns of the diseases of interest identified in this study were previously known. Regarding the diseases of interest, these associations can be classified as comorbidities, early manifestations initially diagnosed as something else, diseases attributable to or that cause the disease of interest, or caused by pharmacological treatment. International Classification of Diseases, Tenth Revision chapters that were most associated with other chapters included L Diseases of the skin and subcutaneous tissue. In the age-stratified and gender-stratified networks, clusters with the highest numbers of International Classification of Diseases, Tenth Revision codes included I Diseases of the circulatory system and F Mental and behavioral disorders.

Conclusions Our findings reinforce established associations between NCDs and underline the importance of comprehensive NCD care.

1 Introduction

Non-communicable diseases (NCDs), which include cardiovascular diseases, cancer, diabetes mellitus, and chronic lung diseases, are leading causes of global mortality [1]. In 2016, NCDs caused 40.5 million deaths—71% of the global total [2]. Cardiovascular diseases such as stroke and ischemic heart disease accounted for 44% of these NCD deaths.

Non-communicable diseases are also a major cause of morbidity and disability. Mental health disorders such as

Key Points

Reducing the non-communicable disease (NCD) burden in the aging Japanese population depends on a better understanding of temporal relationships between different NCDs.

This study identified associations between NCDs and temporal patterns of NCDs in Japan using data from a large medical claims database.

Our findings reinforced established associations between NCDs and underline the importance of comprehensive NCD care.

✉ Shingo Higa
Shingo.Higa@viatris.com

¹ Viatris Pharmaceuticals Japan Inc., Tokyo, Japan

² MinaCare Co., Ltd., Tokyo, Japan

³ Clinical Study Support, Inc., Nagoya, Japan

anxiety and depression, musculoskeletal disorders, mobility impairment, and chronic pain place a significant burden on patients and their families [3–5]. As well as this burden on individuals, NCDs are associated with a considerable societal burden, causing loss of productivity and absenteeism in working-age individuals. The direct and indirect costs of NCDs have been estimated at USD 95 trillion and are increasing [6].

The situation in Japan generally mirrors that of other high-income countries. Non-communicable diseases accounted for 82% of all deaths in 2016, with cancer accounting for 30% of NCD deaths and cardiovascular diseases 27% [7]. However, mortality from many NCDs has declined or plateaued over recent decades [8]. Similarly, for many NCDs the disease burden, measured as disability-adjusted life-years, decreased after 1990 and has stabilized since 2005 [8]. Nonetheless, Japan is one of the fastest aging countries in the world. Japan's population was the world's most elderly in 2017 (when 33% of population were aged 60 years or over) and is projected to remain so until 2050, when an estimated 42% of the population will be aged 60 years or over [9]. Given that the overall NCD prevalence increases with age, the NCD burden resulting from this aging population is an urgent challenge for public health and medical practice in Japan.

Non-communicable diseases are often interrelated, frequently occurring in the same patients. For example, people with diabetes commonly have comorbid microvascular or macrovascular diseases [10, 11]. A network analysis is an analytical method that can be used to explore the complex relationships among diseases and to provide insight into disease progression pathways [12–14]. It has been used to explore disease associations and progression with large datasets obtained from healthcare claims databases in Europe, North America, and South Korea [14–16]. However, no such studies have been conducted to investigate disease associations and progression of NCDs in Japan.

To develop strategies for reducing the considerable NCD burden in Japan, it is important to understand NCD disease progression pathways in the Japanese population. This will facilitate appropriate medical intervention at an early stage, potentially reducing death and disability caused by NCDs.

The aim of this study was to identify associations between NCDs and temporal patterns of NCDs in Japanese working individuals and their dependents using data from a large medical claims database by a network analysis. This population was selected because medical intervention at younger ages will help decrease the burden of NCDs.

2 Methods

2.1 Source Data and Study Population

This study was based on a database managed by MinaCare Co., Ltd. (Tokyo, Japan). The database includes data for health check-ups and employment-based health insurance claims for medical and pharmaceutical treatment for the period since 2010. Data relate to individuals working for retailers, manufacturers, and in the food, information, transportation, and energy industries (and their dependents). The age range covered by the database is 0–75 years and the median follow-up period for individuals included in the database is < 3 years.

Japan has a universal health insurance system, where all citizens are covered by one of the following public health insurance schemes: health insurance association, mutual aid association, national health insurance, and late-stage medical insurance for the elderly. This study utilized the MinaCare database, which is a health insurance association database.

As of March 2019, medical and pharmaceutical claims data from approximately 5.5 million individuals and health check-up data from approximately 2.3 million individuals were available in the MinaCare database. Data were extracted for all 4,297,827 individuals (the total number of patients in the dataset) with at least one medical or pharmaceutical insurance claim in the period 1 April, 2010 to 31 March, 2018. Whilst no specific inclusion or exclusion criteria were applied, the following International Classification of Diseases, Tenth Revision (ICD-10) codes are not directly related to NCDs and were excluded from the analysis: certain infectious and parasitic diseases (A00–B99); symptoms, signs and abnormal clinical and laboratory findings, not elsewhere classified (R00–R99); injury, poisoning and certain other consequences of external causes (S00–T88); external causes of morbidity (V00–Y99); and factors influencing health status and contact with health services (Z00–Z99). Moreover, we did not restrict the inclusion of disease codes to purely chronic conditions because of the potential risk of misdiagnosing symptoms prior to the diagnosis of chronic conditions.

For each individual, the index date was defined as the date of the earliest recorded diagnosis during the study period, and all diagnoses were included in the analysis.

2.2 Diagnosis Codes

We analyzed the occurrence of NCD diagnoses, identified by ICD-10 codes within the claims data. International Classification of Diseases, Tenth Revision codes were rounded to three digits (the first three digits indicate the main

diagnosis). For example, heart failure is designated by the code “I50”. For each individual identified as having at least one NCD based on the ICD-10 codes included in the claims data, only the first occurrence of each ICD-10 code on or after the index date was considered in the analysis.

2.3 Associations Between NCDs

Associations between pairs of NCDs (three-digit ICD-10 codes) were assessed by calculating risk ratios (RRs) using the following formula [14]:

$$RR_{ij} = \frac{C_{ij}/N}{C_i C_j - N^2},$$

where C_i and C_j are the numbers of individuals with diseases i and j , respectively, in the study population, C_{ij} is the number of individuals affected by both diseases, and N is the total number of individuals in the study population. Ninety-five percent confidence intervals were estimated using the following formula [17]:

$$[RR_{ij} \times e^{-1.96\sigma_{ij}}, RR_{ij} \times e^{1.96\sigma_{ij}}],$$

where

$$\sigma_{ij} = \frac{1}{C_{ij}} + \frac{1}{C_i C_j} - \frac{1}{N} - \frac{1}{N^2}.$$

2.4 Overall Network

Pairs of ICD-10 codes with RR point estimates > 15 and where the 95% confidence interval did not include 1 were included in the overall network analysis. This threshold was set in order to obtain a good compromise between sparsity and density of the network. The threshold was determined using the number of nodes in the largest and second largest components in the network. In network theory, a network undergoes a transition from the phase of connectivity to the phase of disconnectivity. Concretely, the change fraction of nodes and links in the largest and the second largest components was plotted, and the threshold was visually determined. The disease association network was displayed graphically, with the width of the edges indicating the strengths of the associations and the sizes of the nodes indicating disease prevalence. Percentages of network links in each ICD-10 chapter (self-flow) and percentages of links connecting one ICD-10 chapter to another ICD-10 chapter (outflow) were calculated. Similarly, for each ICD-10 code the percentage of links with other ICD-10 chapters and the sum of their corresponding RR values were calculated.

It is important to note that in this study the term risk ratio is also often called the observed-to-expected ratio, which is equivalent to the point-wise mutual information, and is not the traditional RR used in epidemiology.

2.5 Stratification by Sex and Age

Disease networks were constructed for strata based on sex and age at the index date in a similar manner as for the overall network. The age strata were 18 to < 35 years and 35 to < 65 years. Few individuals were aged ≥ 65 years and the data for subjects aged ≥ 65 years were pooled as the third age strata.

For each network, clustering was performed with the Infomap algorithm [16], which uses a community detection method to identify communities or clusters of nodes (ICD-10 codes in the present case) that are highly connected with each other. Nodes in a community are more likely to connect to other nodes in the same community rather than to nodes in other communities.

2.6 Diseases of Interest

We further defined eight diseases of interest—individual 3-digit ICD-10 codes or groups of codes selected to include major contributors to the global disability-adjusted life-year-based disease burden (Table 1) [8, 18]. For grouped diagnoses, RRs were estimated after grouping.

2.7 Temporal Associations

To assess the strengths of the associations between pairs of diagnoses (three-digit ICD-10 codes) in a time-dependent manner, RRs were calculated using the following formula [16]:

$$RR_{i \rightarrow j} = \frac{a \times N}{b \times c},$$

where a is the number of pairs with the i th and j th diseases, with the i th disease as the earlier diagnosis, b is the number of pairs with the i th disease as the earlier diagnosis, c is the number of pairs with the j th disease as the later diagnosis, and N is the total number of pairs. Pairs of three-digit ICD-10 codes with an RR point estimate ≥ 5 and where the 95% confidence interval did not include 1 were identified.

2.8 Temporal Patterns

An analysis of temporal patterns of three-digit ICD-10 codes was conducted based on the methods of Giannoula et al.

[15]. For temporal associations with an RR of ≥ 5 , based on a 5% patient sample, disease record sequence was extracted for each patient in the sample. Sequences longer than ten were truncated to remove rare trajectories. Then, among obtained sequences, all possible disease subsequence patterns with length ≥ 3 were identified. Subsequence patterns common to different patterns were then identified by comparing all subsequence patterns among all patients. Patterns with length ≥ 3 and present in at least ten patients were analyzed, and the number of patients with the different patterns was determined.

2.9 Statistical Software

Statistical analyses were conducted using R version 3.5.3 (R Foundation, Vienna, Austria), Python version 3.7 (Python Software Foundation, Wilmington, DE, USA), and SAS version 9.4 (SAS Institute Inc., Cary, NC, USA).

3 Results

3.1 Demographics

The analysis used data from 4,200,254 individuals, who had at least one documented NCD (Table 1). 58.57% of individuals were female. Age at the index date was < 18 years in 21.86% of individuals, 18 to < 35 years in 32.85%, 35 to < 65 years in 43.40%, and ≥ 65 years in 1.88%. The mean follow-up time was 2.57 years. The median, first, and third quartiles of follow-up time were 2, 0.67, and 4.00 years, respectively.

Table 1 Diagnoses of interest related to global disease burden

Number	ICD-10 code	ICD-10 name
1	M05	Seropositive rheumatoid arthritis
	M06	Other rheumatoid arthritis
2	M17	Gonarthrosis
3	M16	Coxarthrosis
4	M54	Dorsalgia
5	G98	Other disorders of nervous system, not elsewhere classified
6	F32	Depressive episode
	F33	Recurrent depressive disorder
7	F41	Other anxiety disorders
8	I50	Heart failure

Selected NCDs with high burden of morbidity and mortality [8, 18]

3.2 NCD Associations in the Overall Network

The ICD-10 code-based disease associations in the overall network are shown in Fig. 1. The RRs for the diseases having the 30 highest RR values, showing the diseases with the highest associations, are displayed in Table 1 of the Electronic Supplementary Material (ESM). The ICD-10 codes accounting for the greatest proportions of associations were I10 Essential (primary) hypertension (6.05%), E28 Ovarian dysfunction (5.56%), and F32 Depressive episode (4.11%).

3.3 ICD-10 Chapter-Based Disease Associations in the Overall Network

There were 601,583 individuals included in at least one significant association (RR) between different disease chapters. Sums of RRs were especially high for ICD-10 chapters P Certain conditions originating in the perinatal period (29,787.9) and O Pregnancy, childbirth and the puerperium (25,651.3) (Table 2). The ICD-10 chapters for which self-flow (associations within the same chapter) accounted for the highest percentages of summed RR values were H Diseases of the eye and adnexa, Diseases of the ear and mastoid process (85.5%), O Pregnancy, childbirth and the puerperium (75.0%), and P Certain conditions originating in the perinatal period (67.2%).

The chapters with the highest percentages of outflow (links with other chapters) were L Diseases of the skin and subcutaneous tissue (90.7%; 93.5% of the sum of RR values), G Diseases of the nervous system (79.5%; 87.9%), and E Endocrine, nutritional and metabolic diseases (75.7%; 81.0%) (Table 2). Outflow for individual ICD-10 codes is shown in Table 2 of the ESM.

3.4 Clusters by Sex and Age Group (Individuals Aged 18 to < 35 Years and 35 to < 65 Years) and Their Correlations

The top three clusters with the highest numbers of ICD-10 codes for individuals stratified by sex and age are shown in Fig. 1a, b of the ESM. For male individuals aged 18 to < 35 years, the clusters with the highest numbers of ICD-10 codes mainly contain codes relating to chapter I Diseases of the circulatory system, and centered on I10 Essential (primary) hypertension; chapter F Mental and behavioral disorders; and chapter N Diseases of the genitourinary system, and centered on N41 Inflammatory diseases of prostate and N30 Cystitis. For male individuals aged 35 to < 65 years, the clusters with the highest numbers of ICD-10 codes mainly consist of chapter N Diseases of the genitourinary system; chapter I Disease of the circulatory systems, centered on I50 Heart failure; and chapter H Diseases of the eye and adnexa,

Table 2 ICD-10 chapter-based disease associations in the overall network (self-flow/outflow)

ICD-10 chapter	(A) Number of ICD-10 codes within chapter	(B) Number of links in network	(C) Number of links with other chapters	(D) % of links in the network	(E) % of links with other chapters	(F) Summed RRs for links in network	(G) Summed RRs with other chapters	(H) % of summed RRs in the network	(I) Out-flow as a % of summed RRs in the network
C Malignant neoplasms	32	149	117	8.2	78.5	13,249.9	8,436.2	12.4	63.7
D In situ neoplasms, benign neoplasms, neoplasms of uncertain or unknown behavior, diseases of the blood and blood-forming organs and certain disorders involving the immune mechanism	36	163	137	9.0	84.0	9382	7,271.6	8.8	77.5
E Endocrine, nutritional and metabolic diseases	40	230	174	12.6	75.7	12,130.7	9,820.1	11.3	81.0
F Mental and behavioral disorders	33	118	58	6.5	49.2	6,632	3,931.5	6.2	59.3
G Diseases of the nervous system	22	88	70	4.8	79.5	5,643	4,960.2	5.3	87.9
H Diseases of the eye and adnexa, Diseases of the ear and mastoid process	25	38	5	2.1	13.2	1,207.5	175.5	1.1	14.5
I Diseases of the circulatory system	43	218	89	12.0	40.8	11,062.9	6,955.7	10.3	62.9
J Diseases of the respiratory system	23	64	42	3.5	65.6	2,649.7	1,981.4	2.5	74.8
K Diseases of the digestive system	33	100	59	5.5	59.0	7,012.9	5,027.6	6.6	71.7

Table 2 (continued)

ICD-10 chapter	(A) Number of ICD-10 codes within chapter	(B) Number of links in network	(C) Number of links with other chapters	(D) % of links in the network	(E) % of links with other chapters	(F) Summed RRs for links in the network	(G) Summed RRs with other chapters	(H) % of summed RRs in the network	(I) Out-flow as a % of summed RRs in the network
L Diseases of the skin and subcutaneous tissue	13	54	49	3.0	90.7	1,605.7	1,501.6	1.5	93.5
M Diseases of the musculoskeletal system and connective tissue	26	76	46	4.2	60.5	4,307.6	2,796.5	4.0	64.9
N Diseases of the genitourinary system	52	277	206	15.2	74.4	11,630.8	8,432.2	10.9	72.5
O Pregnancy, childbirth and the puerperium	49	664	189	36.5	28.5	25,651.3	6,417.2	24.0	25.0
P Certain conditions originating in the perinatal period	25	253	167	13.9	66.0	29,787.9	9,777.1	27.9	32.8
Q Congenital malformations, deformations and chromosomal abnormalities	17	59	54	3.2	91.5	6,023.5	4,638.3	5.6	77.0

D) = B/total number of links in the network $\times 100$; E) = C/B) $\times 100$; H) = F/overall sum of RRs for disease links in the network $\times 100$; I) = G/F) $\times 100$
RR risk ratio

Table 3 Temporal associations of diagnoses of interest with other diagnoses in time sequence patterns of observation (1 → 2 → 3)

ICD-10 code of disease associated with main disease of interest (1)		Risk ratio	ICD-10 code and main disease of interest (2)		ICD-10 code of disease associated with main disease of interest (3)		Risk ratio
J84	Other interstitial pulmonary diseases	12.10	M05	Seropositive rheumatoid arthritis	J84	Other interstitial pulmonary diseases	27.29
-			M06	Other rheumatoid arthritis	D89	Other disorders involving the immune mechanism, not elsewhere classified	19.59
-					E53	Deficiency of other B group vitamins	16.68
M15	Polyarthrosis	5.90			-		
M19	Other arthrosis	5.29			-		
M32	Systemic lupus erythematosus (SLE)	13.65			M32	Systemic lupus erythematosus (SLE)	17.55
M35	Other systemic involvement of connective tissue	12.90			M35	Other systemic involvement of connective tissue	13.69
-					M81	Osteoporosis without pathological fracture	7.31
-							
-			M17	Gonarthrosis	M87	Osteonecrosis	9.00
M16	Coxarthrosis	5.25			M16	Coxarthrosis	5.06
-							
-			M16	Coxarthrosis	M87	Osteonecrosis	24.21
Q65	Congenital deformities of hip	11.33			Q65	Congenital deformities of hip	19.32
M48	Other spondylopathies	6.60			M48	Other spondylopathies	7.61
-					M81	Osteoporosis without pathological fracture	5.45
M17	Gonarthrosis	5.06			M17	Gonarthrosis	5.25
M43	Other deforming dorsopathies	5.34			-		

Table 3 (continued)

-			M54	Dorsalgia	-		
M48	Other spondylopathies	8.57	G98	Other disorders of nervous system not elsewhere classified	M48	Other spondylopathies	9.59
G64	Other disorders of peripheral nervous system	6.59			G64	Other disorders of peripheral nervous system	6.55
G57	Mononeuropathies of lower limb	5.94			G57	Mononeuropathies of lower limb	6.96
M50	Cervical disc disorders	5.82			M50	Cervical disc disorders	6.21
G56	Mononeuropathies of upper limb	5.14			G56	Mononeuropathies of upper limb	5.08
G54	Nerve root and plexus disorders	5.65			-		
M51	Other intervertebral disc disorders	5.11			-		
-			F32	Depressive episode	F21	Schizotypal disorder	15.25
F31	Bipolar affective disorder	6.31	F33	Recurrent depressive disorder	F31	Bipolar disorder	13.94
F25	Schizoaffective disorders	6.60			F25	Schizoaffective disorders	13.43
F22	Persistent delusional disorders	6.31			F22	Persistent delusional disorders	13.52
F20	Schizophrenia	6.38			F20	Schizophrenia	12.60
G21	Secondary parkinsonism	5.66			G21	Secondary parkinsonism	12.32
F30	Manic episode	5.91			F30	Manic episode	12.09
F60	Specific personality disorders	7.85			F60	Specific personality disorders	10.93
F43	Reaction to severe stress, and adjustment disorders	10.46			F43	Reaction to severe stress, and adjustment disorders	7.07
F40	Phobic anxiety disorders	10.29			F40	Phobic anxiety disorders	9.83
F42	Obsessive-compulsive disorder	9.83			F42	Obsessive-compulsive disorder	9.17
F44	Dissociative [conversion] disorders	5.42			F44	Dissociative [conversion] disorders	8.92

Table 3 (continued)

-					F23	Acute and transient psychotic disorders	8.62
-					G20	Parkinson disease	8.66
-					G40	Epilepsy	7.19
F99	Mental disorder, not otherwise specified	7.19			F99	Mental disorder, not otherwise specified	5.96
F48	Other neurotic disorders	6.39			-		
F41	Other anxiety disorders	6.34			-		
F34	Persistent mood [affective] disorders	6.33			F34	Persistent mood [affective] disorders	5.35
G47	Sleep disorders	5.90			G47	Sleep disorders	6.17
-					F51	Nonorganic sleep disorders	6.12
-					F90	Hyperkinetic disorders	5.99
-					G25	Other extrapyramidal and movement disorders	5.84
-							
-			F41	Other anxiety disorders	F40	Phobic anxiety disorders	7.15
-					F31	Bipolar disorder	6.61
-					F20	Schizophrenia	6.60
-					F60	Specific personality disorders	6.61
-					F32	Depressive episode	6.34
-					F34	Persistent mood [affective] disorders	6.15
-					G21	Secondary parkinsonism	6.07
-					F42	Obsessive-compulsive disorder	6.04
-					F22	Persistent delusional disorders	5.96
-					F30	Manic episode	5.77
-					F25	Schizoaffective disorders	5.54

Table 3 (continued)

-					F44	Dissociative [conversion] disorders	5.10
I42	Cardiomyopathy	18.43	I50	Heart failure	I42	Cardiomyopathy	26.73
I27	Other pulmonary heart diseases	11.81			I27	Other pulmonary heart diseases	25.17
I21	Acute myocardial infarction	19.88			I21	Acute myocardial infarction	14.68
I48	Atrial fibrillation and flutter	19.81			I48	Atrial fibrillation and flutter	16.73
I25	Chronic ischemic heart disease	12.61			I25	Chronic ischemic heart disease	16.46
I71	Aortic aneurysm and dissection	15.19			I71	Aortic aneurysm and dissection	12.53
I35	Nonrheumatic aortic valve disorders	14.04			I35	Nonrheumatic aortic valve disorders	14.15
Q21	Congenital malformations of cardiac septa	9.42			Q21	Congenital malformations of cardiac septa	11.68
I34	Nonrheumatic mitral valve disorders	11.53			I34	Nonrheumatic mitral valve disorders	12.26
I38	Endocarditis, valve unspecified	10.98			I38	Endocarditis, valve unspecified	12.07
-					N17	Acute kidney failure	11.30
I44	Atrioventricular and left bundle-branch block	10.99			-		
I20	Angina pectoris	10.37			I20	Angina pectoris	9.21
I47	Paroxysmal tachycardia	9.48			I47	Paroxysmal tachycardia	10.79
-					J90	Pleural effusion, not elsewhere classified	10.29
I11	Hypertensive heart disease	8.42			I11	Hypertensive heart disease	10.43
I07	Rheumatic tricuspid valve diseases	8.19			I07	Rheumatic tricuspid valve diseases	9.59
N19	Unspecified kidney failure	8.22			N19	Unspecified kidney failure	9.13
I51	Complications and ill-defined descriptions of heart disease	8.30			I51	Complications and ill-defined descriptions of heart disease	8.38

Table 3 (continued)

N18	Chronic kidney disease (CKD)	7.41		N18	Chronic kidney disease (CKD)	7.96
-				E88	Other metabolic disorders	6.99
-				I74	Arterial embolism and thrombosis	6.66
J84	Other interstitial pulmonary diseases	6.63		J84	Other interstitial pulmonary diseases	6.42
I49	Other cardiac arrhythmias	6.46		-		
Q25	Congenital malformations of great arteries	5.95		-		
E11	Type 2 diabetes mellitus	5.33		E11	Type 2 diabetes mellitus	5.52
I10	Essential (primary) hypertension	5.40		-		
I45	Other conduction disorders	5.62		-		
-				E83	Disorders of mineral metabolism	5.20
-				I69	Sequelae of cerebrovascular disease	5.06
-				D65	Disseminated intravascular coagulation [defibrination syndrome]	5.07
-				D80	Immunodeficiency with predominantly antibody defects	6.75
-				E21	Hyperparathyroidism and other disorders of parathyroid gland	8.96

Associations with risk ratios ≥ 5 are shown. Diagnoses of interest are shown in the middle column. Associations: blue = antecedent; red = subsequent; black = bidirectional.

F31 Bipolar disorder, F25 Schizoaffective disorders, F22 Delusional disorders, F20 Schizophrenia, and F48 Obsessive compulsive disorder, as well as with G21 Secondary parkinsonism (Table 3). It also frequently preceded other psychiatric diseases, including F21 Schizotypal disorder, F23 Acute and transient psychotic disorders, and F90 Hyperkinetic disorders, as well as G20 Parkinson's disease. Similarly, F41 Other anxiety disorders showed unidirectional temporal associations with various other psychiatric diseases, often preceding F31 Bipolar affective disorder, F20 Schizophrenia, F60 Specific personality disorders, and F32 Depressive episode. Both major depressive disorder and anxiety disorders were included together in multiple temporal patterns, many of which also included G47 Sleep disorders or F20 Schizophrenia (Table 3 of the ESM).

Finally, I50 Heart failure had bidirectional temporal associations with various other diseases of the circulatory system, including I20 Angina pectoris, I21 Acute myocardial infarction, I42 Cardiomyopathy, and I48 Atrial fibrillation and flutter, as well as with other chronic diseases such as E11 Type 2 diabetes mellitus and N18 Chronic kidney disease (Table 3). It was also often preceded by I10 Essential (primary) hypertension. Temporal patterns frequently included E78 Disorders of lipoprotein metabolism and other lipidemias (Table 3 of the ESM).

In the study results, some known disease associations of interest were not captured. These included comorbidities of low back pain, comorbidities of depression, and comorbidities of heart failure. Regarding comorbidities of low back

pain (M545) included in dorsalgia (M54), no associations were observed with any psychiatric disorder [19].

Regarding comorbidities of depression (F32, F33), no associations were observed with substance use disorders or substance-related disorders (F10–F19 mental and behavioral disorders due to psychoactive substance use). No associations were also observed with anorexia nervosa or bulimia nervosa (F50 eating disorders) [20]. Regarding comorbidities of heart failure (I50), no associations were observed with hyperuricemia (E790)/gout (M10), anemia (D50–D64), or sleep apnea (G473) [21].

4 Discussion

4.1 Study Background, Overall Results, and Global Relevance

This is the first study to use a network analysis of data from a large medical claims database to identify associations between NCDs and temporal patterns of NCDs in Japanese working individuals and their dependents. Some ICD-10 chapters demonstrated high self-flow in the network, indicating that the diseases they contain are tightly connected to each other. As expected, these chapters include O Pregnancy, childbirth and the puerperium and P Certain conditions originating in the perinatal period, which made major contributions to the overall network in terms of sums of RRs. These findings are partially consistent with a previous study based on ICD-9 codes for medically insured people in

Québec (Canada), which found that a large share of the overall link weight in a coarse 17-node network was accounted for by the category “complications of pregnancy, childbirth, and the puerperium” but not by the category “certain conditions originating in the perinatal period” [22]. In our study, chapters in the network that were highly connected with other chapters (i.e., chapters with high outflow) included L Diseases of the skin and subcutaneous tissue, G Diseases of the nervous system, and E Endocrine, nutritional and metabolic diseases. These chapters represent diseases that have previously been shown to commonly occur as comorbidities of other diseases.

4.2 Results Stratified by Sex and Age

When the sample was stratified by sex and age, clusters with the highest numbers of ICD-10 codes included I Diseases of the circulatory system and F Mental and behavioral disorders. It is thought that consideration of treatment approaches to these major disease groups is important from the perspective of comprehensive NCD care. A recent meta-analysis established that people with severe mental illness are at increased risk of cardiovascular disease [23]. Disease management in people with mental disorders should consider the need for an intervention to reduce the risk of cardiovascular disease.

4.3 Temporal Associations and Patterns for the Diseases of Interest

Many of the temporal associations and temporal patterns of the diseases of interest identified in this study were previously known. These associations can be classified as reflecting comorbidities of the disease of interest; early manifestations of the disease of interest that were initially diagnosed as something else; diseases attributable to or that cause the disease of interest; or comorbidities caused by pharmacological treatment of the disease of interest. For RA, frequent temporal associations and patterns included J84 Other interstitial pulmonary diseases and M35 Other systemic involvement of connective tissue, the latter of which includes Sjogren’s syndrome [24]. These findings are in accord with earlier observations that interstitial pneumonia and other pulmonary diseases are common in patients with RA [25, 26]. It should also be noted that when arthritis and arthralgia symptoms are initially noted, it can be difficult to distinguish RA from other musculoskeletal diseases. This may explain why RA was often preceded by M15 Polyarthrosis or M19 Other arthrosis and had a strong temporal association with M32 Systemic lupus erythematosus. We speculate that some of the other temporal associations and patterns may reflect drug-induced conditions: J84 Other interstitial pulmonary

diseases can be induced by disease-modifying anti-rheumatic drugs [25]; E53 Deficiency of other B group vitamins may be related to the prescription of vitamin B to prevent or treat stomatitis as a side effect of methotrexate [27]; and M81 Osteoporosis without a pathological fracture may result from corticosteroid treatment of RA [28].

Our observation of a temporal association between OA of the knee or hip and osteonecrosis may reflect difficulties with differential diagnoses of these distinct diseases. Other observed temporal associations with OA of the hip represent plausible causes of OA (Q65 Congenital deformities of hip) [29] or comorbid conditions (M48 Other spondylopathies and M43 Other deforming dorsopathies) [29]. Although the relationship between hip OA and osteoporosis is controversial [29], our results indicated a possible temporal association. Careful follow-up for osteoporosis may be warranted in patients with OA of the hip.

Consistent with existing evidence [29], we observed an association between OA of the knee and hip. Misalignment caused by arthritis at one joint can affect other joints. Our finding suggests that attention should be paid to secondary OA in other joints in patients diagnosed with OA in the knee or hip.

Dorsalgia showed no significant temporal associations with other ICD-10 codes where the RR was ≥ 5 . It is frequently difficult to diagnose the cause of non-specific low back pain, but this symptom has been linked to various pathological conditions including physical and psychiatric diseases [37]. The results of this study are consistent with prevailing medical knowledge that clear relationships between low back pain and specific diseases were not established.

The diagnoses found to have temporal associations with G98 Other disorders of nervous system not elsewhere classified tend to involve neuropathic pain. The association with M48 Other spondylopathies is reasonable, as a nationwide Japanese study showed that over 50% of patients with spine-related pain were found to have neuropathic pain [30]. Other identified diagnoses, such as G56 Mononeuropathies of upper limb, G57 Mononeuropathies of lower limb, M50 Cervical disc disorders, and M51 Other intervertebral disc disorders are well-established causes of neuropathic pain [31]. The associations with these diagnoses are thus plausible.

Patients with mood disorders often also have other psychiatric diseases, including panic disorder (included under F41 Other anxiety disorders) and obsessive-compulsive disorder [32]. Our results for major depressive disorder confirm these associations. Further, our observation that major depressive disorder often followed a diagnosis of F41 Other anxiety disorders is consistent with the commonly observed disease course in depression [32]. The observed temporal patterns including major depressive disorder, anxiety

disorders, and sleep disorders reflect a well-established confluence of diagnoses [33].

We found that major depressive disorder often preceded other psychiatric diseases. Patients with psychiatric diseases are often initially diagnosed with depression because they tend to visit their physician when their depressive symptoms are prominent. Other symptoms may subsequently become apparent during the course of treatment, leading to concomitant diagnoses. For example, bipolar disorder is often initially diagnosed as depression [34], with the diagnosis of bipolar disorder only occurring months or years later when a manic episode is observed or otherwise confirmed by the physician. This may be because many patients do not recognize mania as a negative condition and therefore do not report it.

Parkinson's disease, which often leaves patients prone to depression, also showed a strong temporal association with major depressive disorder. However, it is known that in the course of Parkinson's disease many non-motor symptoms, including psychiatric symptoms such as major depressive disorder present in the early disease stage before motor function symptoms, are sometimes misdiagnosed as depression [35]. Conversely, after motor function symptoms from Parkinson's disease manifest, facial muscles that control facial expressions do not move normally, and may be mistaken for depression. Alternatively, the temporal association may reflect parkinsonism resulting from treatment with antipsychotics.

Anxiety disorders are symptoms or comorbid conditions for many psychiatric disorders. For example, it has been reported that attention-deficit hyperactivity disorder and depression frequently present as comorbidities with anxiety [36, 37]. In our analysis, temporal associations with anxiety disorders were all unidirectional, with anxiety preceding the other diagnoses. However, in some individuals, it is feasible that other psychiatric disorders diagnosed subsequent to anxiety may actually have manifested (but were not diagnosed) before anxiety.

Several temporal associations with heart failure involved diseases that cause, exacerbate, or are common comorbidities of heart failure, such as E11 Type 2 diabetes mellitus, I10 Essential (primary) hypertension, I20 Angina pectoris, I42 Cardiomyopathy, and N18 Chronic kidney disease [38]. The association with I10 Essential (primary) hypertension was unidirectional, with hypertension preceding heart failure. This reflects the natural disease progression. However, associations with other diseases that cause heart failure, such as type 2 diabetes, were bidirectional. Given that people with type 2 diabetes and hypertension can develop heart failure, public health initiatives should be implemented to decrease the incidence of heart failure.

In the study results, some known disease associations of interest were not captured, including comorbidities of low

back pain, comorbidities of depression (mental and behavioral disorders due to psychoactive substance use and eating disorders), and comorbidities of heart failure (hyperuricemia, anemia, and sleep apnea) were noted. Some relationships of these diseases and their comorbidities are poorly understood, or there is no specific medication to treat some comorbidities. In other cases, medications for the primary disease itself are for the treatment of complications. It is thought that these comorbid disease names were not entered in the claims data because it was not necessary for medical fee billing.

4.4 Future Recommendations

In order to confirm the findings of this study; and to inform health authorities and medical practitioners to implement these results, in preventing NCDs and providing comprehensive NCD care, it is necessary to conduct additional research. Such research may involve the conduct of clinical trials or prospective observational studies using electronic medical records.

4.5 Limitations

Our findings should be viewed in the context of certain study limitations in bias and generalizability, regarding both the data available in the MinaCare database and the use of a network analysis as a statistical technique. The study population was limited to members of the health insurance society covered by the MinaCare database and the results may have limited generalizability. As such, the demographics and socioeconomic status of the study population does not fully represent the overall Japanese population. For instance, men and people aged older than 65 years were under-represented. Another limitation of the MinaCare database is that it does not include workers in primary industries such as agriculture, fisheries, and forestry [39]. The aforementioned features of the database limit the generalizability of the study population to the general Japanese population in both employment type and geography, with regions where primary industries dominate being under-represented in the analysis. However, the large size and nationwide reach of the MinaCare database mean that the analysis should give a fair reflection of the health status of working individuals and their dependents in Japan employment-based health insurance.

Bias may arise from the fact that some diagnostic codes may not have been accurately entered into the database. For example, non-specific codes assigned as "other" may frequently be used for reimbursement purposes and this misclassification of diseases may have caused bias in this study. However, similar to previous studies, these codes were not excluded in order to capture all diagnoses. It is impossible

to estimate the scale of these potential quality issues and the extent to which they affected the validity of our findings.

There are several methodological limitations with a network analysis. Currently, there is no methodology that can address the issue of censoring or determine causality. Additionally, as we did not focus on any one disease and because each disease has a different latent period, we were not able to identify latent periods for diseases in this study. Moreover, because some diseases are only detected a considerable period after they develop, diagnoses for some individuals may not have been registered in the order in which the diseases actually developed. Finally, some diseases may have been missed because of the short median follow-up period. Findings relating to temporal associations and patterns should be interpreted with these limitations in mind.

5 Conclusions

To our knowledge, this is the first study to investigate associations and temporal patterns of NCDs using Japanese medical claims data, and the first network analysis of NCDs in Japan. Our findings reinforce established disease associations and disease progression pathways of NCDs. This study underlines the importance of preventing NCDs and of providing comprehensive NCD care. In order to influence health policy change and inform medical practitioners, to reduce the population burden of NCDs, it may be necessary to conduct further research to confirm the results of this study. Clinical trials and prospective behavioral studies using electronic medical records are suggested as suitable methods to help influence such change.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s40801-022-00310-w>.

Declarations

Funding This work was sponsored and funded by Viatrix Pharmaceuticals Japan Inc., Tokyo, Japan. The medical writing was supported by Robert Phillips at Clinical Study Support, Inc., Nagoya, Japan and funded by Viatrix Pharmaceuticals Japan Inc.

Conflict of interest The authors declare the following financial interests/personal relationships. SH, KN, YK, CS, SM, and YA are employees of Viatrix Pharmaceuticals Japan Inc., Tokyo, Japan. TL is an employee of Clinical Study Support, Inc., Nagoya, Japan and was involved in the study under contract with Viatrix Pharmaceuticals Japan Inc. YY is the President of MinaCare Co., Ltd., Tokyo, Japan that received fees for data usage from Viatrix Pharmaceuticals Japan Inc.

Ethics approval The study was conducted in accordance with legal and regulatory requirements, including data protection laws. It used data obtained in an anonymized format and containing no personal information. Therefore, informed consent and ethics committee approval were

not required because studies using only anonymized data are outside the scope of the Japanese Government's "Ethical Guidelines for Medical and Health Research Involving Human Subjects."

Consent to participate Not applicable.

Consent for publication Not applicable.

Availability of data and material The MinaCare data are proprietary to MinaCare, Co., Ltd. and not publicly available for research purposes. Researchers who would like to access the data for research purposes should contact Dr. Yuji Yamamoto (mc_info@minacare.co.jp) to make a data use agreement and pay a data availability fee.

Code availability Not applicable.

Authors' contributions All authors made substantial contributions to all of the following: (1) the conception and design of the study, acquisition of data, or analysis and interpretation of data; (2) drafting the article or revising it critically for important intellectual content; and (3) final approval of the version to be submitted and meeting the ICMJE authorship criteria.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License, which permits any non-commercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc/4.0/>.

References

1. World Health Organization. Noncommunicable diseases. 2021. https://www.who.int/health-topics/noncommunicable-diseases#tab=tab_1. Accessed 19 Nov 2021.
2. World Health Organization. Total NCD mortality. 2021. <https://www.who.int/data/gho/data/themes/topics/indicator-groups/indicator-group-details/GHO/total-ncd-mortality>. Accessed 19 Nov 2021.
3. Bloom DE, Cafiero E, Jané-Llopis E, et al. The global economic burden of noncommunicable diseases. 2011. https://www3.weforum.org/docs/WEF_Harvard_HE_GlobalEconomicBurdenNonCommunicableDiseases_2011.pdf. Accessed 19 Nov 2021.
4. Goldberg DS, McGee SJ. Pain as a global public health priority. *BMC Public Health*. 2011;11:770.
5. Lisy K, Campbell JM, Tufanaru C, et al. The prevalence of disability among people with cancer, cardiovascular disease, chronic respiratory disease and/or diabetes: a systematic review. *Int J Evid Based Healthc*. 2018;16(3):154–66.
6. Chen S, Kuhn M, Prettner K, Bloom DE. The macroeconomic burden of noncommunicable diseases in the United States: estimates and projections. *PLoS ONE*. 2018;13(11): e0206702.
7. World Health Organization. Noncommunicable diseases (NCD) country profiles. 2018. https://www.who.int/nmh/countries/2018/jpn_en.pdf?ua=1. Accessed 19 Nov 2021.

8. Nomura S, Sakamoto H, Glenn S, et al. Population health and regional variations of disease burden in Japan, 1990–2015: a systematic subnational analysis for the Global Burden of Disease Study 2015. *Lancet*. 2017;390(10101):1521–38.
9. United Nations. World population ageing 2017: highlights. 2017. https://www.un.org/en/development/desa/population/publications/pdf/ageing/WPA2017_Highlights.pdf. Accessed 19 Nov 2021.
10. Long AN, Dagogo-Jack S. Comorbidities of diabetes and hypertension: mechanisms and approach to target organ protection. *J Clin Hypertens (Greenwich)*. 2011;13(4):244–51.
11. Kerr EA, Heisler M, Krein SL, et al. Beyond comorbidity counts: how do comorbidity type and severity influence diabetes patients' treatment priorities and self-management? *J Gen Intern Med*. 2007;22(12):1635–40.
12. Rual JF, Venkatesan K, Hao T, et al. Towards a proteome-scale map of the human protein-protein interaction network. *Nature*. 2005;437(7062):1173–8.
13. Goh KI, Cusick ME, Valle D, et al. The human disease network. *Proc Natl Acad Sci USA*. 2007;104(21):8685–90.
14. Hidalgo CA, Blumm N, Barabási AL, Christakis NA. A dynamic network approach for the study of human phenotypes. *PLoS Comput Biol*. 2009;5(4): e1000353.
15. Giannoula A, Gutierrez-Sacristán A, Bravo Á, et al. Identifying temporal patterns in patient disease trajectories using dynamic time warping: a population-based study. *Sci Rep*. 2018;8(1):4216.
16. Jeong E, Ko K, Oh S, Han HW. Network-based analysis of diagnosis progression patterns using claims data. *Sci Rep*. 2017;7(1):15561.
17. Katz D, Baptista J, Azen SP, Pike MC. Obtaining confidence intervals for the risk ratio in cohort studies. *Biometrics*. 1978;34(3):469–74.
18. GBD 2017 Disease and Injury Incidence and Prevalence Collaborators. Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. *Lancet*. 2018;392(10159):1789–858. [https://doi.org/10.1016/S0140-6736\(18\)32279-7](https://doi.org/10.1016/S0140-6736(18)32279-7) (Epub 2018 Nov 8. Erratum in: *Lancet*. 2019 Jun 22;393(10190):e44. PMID: 30496104; PMCID: PMC6227754).
19. Japanese Orthopaedic Association. Clinical practice guidelines on the management of low back pain [in Japanese]. 2nd ed. Tokyo: Nankodo; 2019.
20. Takahashi S, Ohno Y, Someya T, Kamba S, Ozaki N, Mimura M, et al. Diagnostic and statistical manual of mental disorders (DSM-5) [in Japanese]. Tokyo: Igaku-Shoin; 2014.
21. The Japanese Circulation Society, Japanese Heart Failure Society. JCS 2017/JHFS 2017 guidelines for diagnosis and treatment of acute and chronic heart failure (in Japanese). https://www.j-circ.or.jp/cms/wp-content/uploads/2017/06/JCS2017_tsutsui_h.pdf. Accessed 6 Apr 2022.
22. Fotouhi B, Momeni N, Riolo MA, Buckeridge DL. Statistical methods for constructing disease comorbidity networks from longitudinal inpatient data. *Appl Netw Sci*. 2018;3(1):46.
23. Correll CU, Solmi M, Veronese N, et al. Prevalence, incidence and mortality from cardiovascular disease in patients with pooled and specific severe mental illness: a large-scale meta-analysis of 3,211,768 patients and 113,383,368 controls. *World Psychiatry*. 2017;16(2):163–80.
24. Grant-in-Aid for Scientific Research on Health and Labor Policy Research Project on Intractable Diseases and Research Group on Autoimmune Disease. Sjogren's syndrome clinical practice guidelines, 2017 [in Japanese]. 2017. <https://minds.jcqhc.or.jp/docs/minds/Sjogren's-syndrome/Sjogren's-syndrome.pdf>. Accessed 19 Nov 2021.
25. Roubille C, Haraoui B. Interstitial lung diseases induced or exacerbated by DMARDS and biologic agents in rheumatoid arthritis: a systematic literature review. *Semin Arthritis Rheum*. 2014;43(5):613–26.
26. Nakajima A, Inoue E, Shimizu Y, et al. Presence of comorbidity affects both treatment strategies and outcomes in disease activity, physical function, and quality of life in patients with rheumatoid arthritis. *Clin Rheumatol*. 2015;34(3):441–9.
27. Gono T. Methotrexate (MTX) practice guideline for the treatment of rheumatoid arthritis revised 2016 (simplified version) [in Japanese]. *J Jpn Soc Intern Med*. 2017;106(7):1433–9.
28. Yeap SS, Hosking DJ. Management of corticosteroid-induced osteoporosis. *Rheumatology (Oxford)*. 2002;41(10):1088–94.
29. Japan Orthopaedic Association. Clinical practice guideline on the management of osteoarthritis of the hip [in Japanese]. 2nd ed. Tokyo: Nankodo Co Ltd; 2016.
30. Yamashita T, Takahashi K, Yonenobu K, Kikuchi S. Prevalence of neuropathic pain in cases with chronic pain related to spinal disorders. *J Orthop Sci*. 2014;19(1):15–21.
31. The Japan Society of Pain Clinicians. Guidelines for the pharmacotherapy of neuropathic pain [in Japanese]. Revised 2nd edition. Tokyo: Shinko Trading Co Ltd. Publication Department; 2016.
32. Iga J, Uchiyama M, Omori T, et al. Japanese society for the study of depression treatment guidelines II. Major depressive disorder 2016 [in Japanese]. *Intern Med*. 2020;62(5):542–9.
33. van Mill JG, Hoogendijk WJ, Vogelzangs N, et al. Insomnia and sleep duration in a large cohort of patients with major depressive disorder and anxiety disorders. *J Clin Psychiatry*. 2010;71(3):239–46.
34. Watanabe K, Harada E, Inoue T, et al. Perceptions and impact of bipolar disorder in Japan: results of an internet survey. *Neuropsychiatr Dis Treat*. 2016;12:2981–7.
35. Kalia LV, Lang AE. Parkinson's disease. *Lancet*. 2015;386(9996):896–912.
36. Kessler RC, Adler L, Barkley R, et al. The prevalence and correlates of adult ADHD in the United States: results from the National Comorbidity Survey Replication. *Am J Psychiatry*. 2006;163(4):716–23.
37. Lamers F, van Oppen P, Comijs HC, et al. Comorbidity patterns of anxiety and depressive disorders in a large cohort study: the Netherlands Study of Depression and Anxiety (NESDA). *J Clin Psychiatry*. 2011;72(3):341–8.
38. Tsutsui H. Guidelines for diagnosis and treatment of acute and chronic heart failure (JCS 2017/JHFS 2017) [in Japanese]. *J Jpn Soc Intern Med*. 2019;108(5):978–85.
39. Shima D, Ii Y, Yamamoto Y, et al. A retrospective, cross-sectional study of real-world values of cardiovascular risk factors using a healthcare database in Japan. *BMC Cardiovasc Disord*. 2014;14:120.