


# Multimorbidity and mortality: A data science perspective

Journal of Multimorbidity and Comorbidity  
Volume 12: 1–20  
© The Author(s) 2022  
Article reuse guidelines:  
[sagepub.com/journals-permissions](https://sagepub.com/journals-permissions)  
DOI: 10.1177/26335565221105431  
[journals.sagepub.com/home/cob](https://journals.sagepub.com/home/cob)  


Kien Wei Siah<sup>1,2</sup> , Chi Heem Wong<sup>1,2,3</sup> , Jerry Gupta<sup>4</sup> and Andrew W Lo<sup>1,2,3,5</sup> 

## Abstract

**Background:** With multimorbidity becoming the norm rather than the exception, the management of multiple chronic diseases is a major challenge facing healthcare systems worldwide.

**Methods:** Using a large, nationally representative database of electronic medical records from the United Kingdom spanning the years 2005–2016 and consisting over 4.5 million patients, we apply statistical methods and network analysis to identify comorbid pairs and triads of diseases and identify clusters of chronic conditions across different demographic groups. Unlike many previous studies, which generally adopt cross-sectional designs based on single snapshots of closed cohorts, we adopt a longitudinal approach to examine temporal changes in the patterns of multimorbidity. In addition, we perform survival analysis to examine the impact of multimorbidity on mortality.

**Results:** The proportion of the population with multimorbidity has increased by approximately 2.5 percentage points over the last decade, with more than 17% having at least two chronic morbidities. We find that the prevalence and the severity of multimorbidity, as quantified by the number of co-occurring chronic conditions, increase progressively with age. Stratifying by socioeconomic status, we find that people living in more deprived areas are more likely to be multimorbid compared to those living in more affluent areas at all ages. The same trend holds consistently for all years in our data. In general, hypertension, diabetes, and respiratory-related diseases demonstrate high in-degree centrality and eigencentality, while cardiac disorders show high out-degree centrality.

**Conclusions:** We use data-driven methods to characterize multimorbidity patterns in different demographic groups and their evolution over the past decade. In addition to a number of strongly associated comorbid pairs (e.g., cardiac-vascular and cardiac-metabolic disorders), we identify three principal clusters: a respiratory cluster, a cardiovascular cluster, and a mixed cardiovascular-renal-metabolic cluster. These are supported by established pathophysiological mechanisms and shared risk factors, and largely confirm and expand on the results of existing studies in the medical literature. Our findings contribute to a more quantitative understanding of the epidemiology of multimorbidity, an important pre-requisite for developing more effective medical care and policy for multimorbid patients.

## Keywords

Multimorbidity, network analysis, survival analysis

Received 17 January 2022; accepted: 15 May 2022

## Introduction

Multimorbidity, defined as the coexistence of two or more chronic medical conditions in an individual patient,<sup>1</sup> is a growing public health concern for healthcare systems worldwide. It has been found to be associated with

<sup>1</sup>Laboratory for Financial Engineering, Sloan School of Management, Massachusetts Institute of Technology, Cambridge, MA, USA

<sup>2</sup>Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA, USA

<sup>3</sup>Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, MA, USA

<sup>4</sup>Digital Catalyst, Swiss Re, Cambridge, MA, USA

<sup>5</sup>Sante Fe Institute, Santa Fe, NM, USA

## Corresponding author:

Andrew W Lo, MIT Sloan School of Management, 100 Main Street, E62–618, Cambridge, MA 02142, USA.

Email: [alo-admin@mit.edu](mailto:alo-admin@mit.edu)



Creative Commons Non Commercial CC BY-NC: This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 License (<https://creativecommons.org/licenses/by-nc/4.0/>) which permits non-commercial use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access pages (<https://us.sagepub.com/en-us/nam/open-access-at-sage>).

adverse health outcomes, including a higher risk of mortality, a lower quality of life, increased utilization of health care, and correspondingly higher healthcare costs.<sup>2–14</sup> It is most prevalent in the elderly population, as organs gradually lose full function with the aging process.<sup>8,15–17</sup> With an increasing life expectancy and an aging population, the number of people with multiple health conditions is set to rise, as is public expenditure on long-term medical care. Unfortunately, current healthcare systems are largely designed to treat single diseases, resulting in the need to use multiple services to manage multimorbidity.<sup>11,18–20</sup> Due to poor coordination and integration in medical care, causing a lack of continuity in treatment, disorders not designated as the primary condition are often undertreated.<sup>21</sup>

In order to align medical care more closely to the needs of patients with multiple health conditions, a better understanding of the epidemiology of multimorbidity in the general population is necessary. Studies have shown that multimorbidity can be present in all age groups, including the pediatric population.<sup>22</sup> In particular, significant attention has been paid to multimorbidity in the elderly, due to its high prevalence in that population. Data sources used range from structured databases (e.g., electronic health records and insurance billing data) to self-administered questionnaires and research interviews. While the former tend to be more reliable, the latter are typically subject to self-reporting bias. Some analyses are based on small sample sizes from selected populations, which likely do not generalize well. Lastly, many studies employ only a narrow range of methods to study multimorbidity patterns (e.g., identifying the most prevalent pairs and triads, or calculating the odds ratio) although some have explored more novel clustering approaches as well (e.g., matrix factorization, association rules, and undirected network analysis).<sup>2,19,20,23–31</sup>

In this paper, we aim to characterize multimorbidity patterns not only in older patients, but also across groups with different demographic and socioeconomic statuses, using a large, nationally representative primary care electronic medical records database. Unlike many previous studies, which generally adopt cross-sectional designs based on single snapshots of closed cohorts, we examine temporal changes in the patterns of multimorbidity across a decade of open-cohort patient data. Among previous longitudinal studies, few have examined disease trajectories.<sup>32</sup> Here, we apply various statistical methods to identify common comorbid pairs and triads of diseases, and use directed and undirected network analysis algorithms to measure temporal multimorbidity progression and identify clusters of chronic conditions. In addition, we analyze the impact of multimorbidity on mortality using survival analysis models.

## Methods

### Data

We use anonymized electronic medical records from The Health Improvement Network (THIN)<sup>33</sup> database for our analysis. The database contains longitudinal patient data collected at primary care clinics throughout the UK, covering approximately 6% of the UK population. The average length of follow-up in the THIN database is around 9 years. We extract demographic information (e.g., date of birth, sex, geographical location, and socioeconomic group), baseline vitals (e.g., smoking and alcohol status), and medical history (e.g., medical condition and date of diagnosis) from patient records between 2005 and 2016. To capture temporal trends in the population, we perform our analysis sequentially on each year of data in the sample period (i.e., one set of results for each year). We categorize the subjects into seven mutually exclusive age groups based on Medical Subject Headings (MeSH) definitions (see [Supplementary Material A](#)).<sup>34</sup> In contrast with studies that use static baseline demographics collected at the beginning of follow-up, we use the point-in-time patient age for our analyses. For example, a patient that is 16 years old in 2005 will be classified as an Adolescent for analyses between 2005 and 2007, and subsequently reclassified as an Adult from 2008 onwards.

Diagnoses are recorded in the THIN database using Read Codes, a coded thesaurus of clinical terms used by the UK National Health Service since 1985.<sup>35</sup> There is no standard method for the selection and definition of morbidities in the literature. After consulting with medical officers and Life & Health (L&H) actuaries at Swiss Re, we identify chronic conditions in the records, that is, diseases that are either permanent, caused by nonreversible pathological alterations, or require long periods of rehabilitation and care,<sup>19,36</sup> and map them to a list of 46 higher level morbidities. Furthermore, we classify the morbidities into 14 System Organ Classes (SOCs) as defined in the Medical Dictionary for Regulatory Activities (MedDRA). (See [Figures 1](#) and [2](#) and [Supplementary Material A](#) for lists of morbidities and classifications.) As in similar studies, we define multimorbidity as the presence of at least 2 of the 46 morbidities in a patient.

### Statistical analysis

We examine the distribution of multimorbidity in relation to age and socioeconomic status, as done in Barnett et al.<sup>11</sup> However, we use the Index of Multiple Deprivation (IMD) as a proxy for socioeconomic status. The IMD is a widely used measure of relative deprivation or poverty of wards and districts in the UK. It is computed using census data as a weighted index of deprivation in seven domains, including income,

Chronic Conditions			
1 Angina	13 Liver-related	25 Melanoma	37 Prostate Cancer
2 Atrial Fibrillation	14 Asthma	26 Cancer-related	38 Testicular Cancer
3 CAD	15 HIV/AIDS	27 Pancreas Cancer	39 Uterus Cancer
4 Cardiac Arrhythmia	16 Multiple Sclerosis	28 Encephalitis	40 COPD
5 Heart Failure	17 Diabetes	29 Parkinson's Disease	41 Respiratory-related
6 HVD	18 Lupus	30 Stroke	42 Aortic Aneurysm
7 MI	19 Breast Cancer	31 Stroke-related	43 Hypertension
8 Cardiac-related	20 Cervix Cancer	32 Dementia	44 Other Aneurysms
9 Thyroid Cancer	21 Colorectal Cancer	33 Bladder Cancer	45 PAD
10 Osophageal Cancer	22 Leukemias	34 Kidney Disease	46 TIA
11 Stomach Cancer	23 Liver Cancer	35 Kidney-related	
12 Liver Disease	24 Lung Cancer	36 Ovarian Cancer	

**Figure 1.** Mapping between index and chronic conditions. CAD: Coronary Artery Disease; HVD: Heart Valve Disorder; MI: Myocardial Infarction; COPD: Chronic Obstructive Pulmonary Disease; PAD: Peripheral Artery Disease; TIA: Transient Ischemic Attack.

MedDRA System Organ Classes	
Card	Cardiac Disorders
En	Endocrine Disorders
GI	Gastrointestinal Disorders
Hep	Hepatobiliary Disorders
Imm	Immune System Disorders
Me	Metabolism and Nutrition Disorders
Mu	Musculoskeletal and Connective Tissue Disorders
Neop	Neoplasms Benign, Malignant, and Unspecified
Nerv	Nervous System Disorders
Ps	Psychiatric Disorders
Ren	Renal and Urinary Disorders
Repr	Reproductive System and Breast Disorders
Res	Respiratory, Thoracic, and Mediastinal Disorders
Vas	Vascular Disorders

**Figure 2.** Abbreviations for MedDRA SOC's used in figures.

employment, education, health, crime, barriers to housing and services, and living environment.<sup>37</sup> (IMD data was available only for a subset of the patients. See [Supplementary Material B](#) for the sample sizes used in this analysis.) We note that the same approach, defining socioeconomic status by the area of residence, has been used in previous studies.<sup>11,38</sup>

For each age group, we also compute the observed prevalence for all individual, pairs, and triplets of morbidities. By the assumptions of probability theory, we expect diseases that are independent to co-occur at a rate close to the product of the observed prevalence of each individual constituent disease (i.e., the expected prevalence). Therefore, by comparing the ratio of the observed prevalence versus the expected prevalence (i.e., the lift), we can identify pairs and triads of diseases that occur together more frequently than expected by chance,

possibly driven by an underlying pathophysiological mechanism. As a second metric, we estimate the odds ratio using logistic regression models to determine the association between each pair of diseases, both without adjustment and adjusted by age, sex, and all other diseases.

Next, we construct multimorbidity networks to study the natural clustering of diseases in the dataset. We consider diseases as nodes with sizes proportional to their observed prevalence. For each pair of diseases, we connect their nodes with an undirected edge weighted by the estimated lift, a measure of the strength of the association between the comorbid pair. This creates a dense network where each node is linked to almost every other node. This density, however, makes visualization and inference difficult. As a pre-processing step for subsequent analysis, we extract the main graph structure by removing edges from the adjacency matrix that are peripheral and relatively unimportant. We prune the edges between nodes that have joint prevalence below the 90th percentile, and keep only the edges that have a lift above 2.0, that is, those edges between pairs that co-occur two times more frequently than expected by chance.<sup>2,39–42</sup>

We compute measures of centrality to identify the most important vertices in the multimorbidity network. In particular, for each node, we compute the degree centrality, which is defined as the number of links incident on a node, a direct measure of the connectivity of a node. In this context, a disease with high degree centrality is important because it often co-occurs with a large number of pathologies. We also estimate the eigenvector centrality, a measure of the transitive influence of nodes. To calculate the eigencentality, each node is assigned a score that is proportional to the sum of the scores of all of its neighbors. Nodes with high eigencentality either have many connections, or are connected to important neighbors. In addition, we compute the

graph clustering coefficient (also known as the transitivity) as a quantitative measure of the network's tendency to aggregate in smaller subgroups. To identify any clusters embedded in the multimorbidity networks, we apply a community detection algorithm based on modularity maximization<sup>43–45</sup> to partition nodes into groups that have dense intra-group connections and sparse inter-group connections. Communities identified in this manner can be interpreted as clusters of diseases that tend to co-occur together.

To gain insight into temporal disease associations, we construct directed multimorbidity networks. We extract from each patient's medical history a sequence of diseases ordered by the time of diagnosis. Using these trajectories, we can derive the probability of any given disease conditional on some prior diagnosis, that is,  $\text{Prob}(\text{Disease B given Disease A})$ . We use these probabilities as weights of the directed edges in the network. As before, we prune the network based on node prevalence and edge weights. Since these connections are directed, we can compute the in-degree and out-degree centralities, defined as the number of edges directed to the node, and the number of edges directed from the node to others, respectively. A node with a high in-degree centrality is often diagnosed following other diseases; a node with a high out-degree centrality often leads to subsequent diagnoses in other diseases. These metrics are useful for understanding disease progression, and any causal or contributory relationships between diseases.

Finally, we examine the association between multimorbidity and mortality by performing predictive survival analysis on the dataset. We use five-year overall survival as the primary outcome variable, and consider in our models a range of features, including demographic group, baseline vitals, baseline medical history, the severity of multimorbidity as quantified by the number of co-occurring chronic conditions, and the presence of any of the top ten most prevalent pairs and triplets of morbidities as observed in the Aged and Elderly age groups. We exclude those subjects aged 65 or less from this part of the analysis, as younger age groups have five-year overall mortality rates close to zero.

We explore three standard methods used in survival modeling—the Cox proportional hazards model,<sup>46</sup> the regularized Cox model, and the accelerated failure time model—and additionally, we apply a nonlinear and non-parametric neural network survival model.<sup>47</sup> For model estimation and validation, we randomly split the original dataset into two disjoint sets, a training set that comprises 70% of the data, and a testing set that comprises the remaining 30%. We use the training set to estimate our models, and keep the testing set as an out-of-sample dataset for performance validation. We use the concordance index (C-index) as the metric for model performance. This metric is commonly used in survival analysis to evaluate its

predictive power.<sup>48</sup> It is a measure of the concordance between orderings of observed survival times and the predicted times or risks. (A C-index of 0.5 corresponds to a random model, while a value of 1.0 corresponds to a perfect model.) We use cross-validation to tune the hyperparameters of the models.

In addition to discriminative power, we assess the calibration of our models by comparing the actual and the predicted survival probabilities at 36, 48, and 60 months of overall survival. For each time cutoff, we divide the test set into quintiles based on the predicted risk scores. We then compute the average predicted score and the true survival probability observed in each of the quintiles. Last, we create calibration plots by plotting the observed probabilities against the predicted probabilities. In the ideal case, the points should lie as close as possible to the diagonal line, which represents perfect calibration.

## Results

### Summary statistics

We summarize the demographic statistics of the study population in [Table 1](#). On average, the dataset consists of approximately 4.6 million patients each year, with an even mix of both sexes in all years. Most of the patient records were collected in England, which makes up the largest part of the population of the United Kingdom. However, the distribution in geographical location has evolved over the years, shifting towards other regions in the country. Over 60% of the patients are in the Adult (19–45 years old) and Middle-Aged (45–65 years old) age groups, as defined by the MeSH classification (see [Supplementary Material A](#)). Approximately 15% are over 65 years old.

The proportion of the population with multimorbidity has increased by approximately 2.5 percentage points over the last decade, with more than 17% of all patients having at least two chronic morbidities in 2016. We find that the prevalence and the severity of multimorbidity increase progressively with age (see [Figure 3](#)). By age 60, approximately half the population has been diagnosed with at least one chronic condition, after which we observe a steep rise in multimorbidity, with close to 1 in 3 patients having at least two morbidities by age 70. Stratifying the prevalence of multimorbidity by IMD, we find that people living in more deprived areas are more likely to be multimorbid compared to those living in more affluent areas at all ages. The same trend holds for all years in our data. (See [Supplementary Material C](#).)

### Individual, pairs, and triplets

We characterize the epidemiology of individual diseases by plotting heat maps of disease prevalence in different age groups. We find that asthma and respiratory conditions have

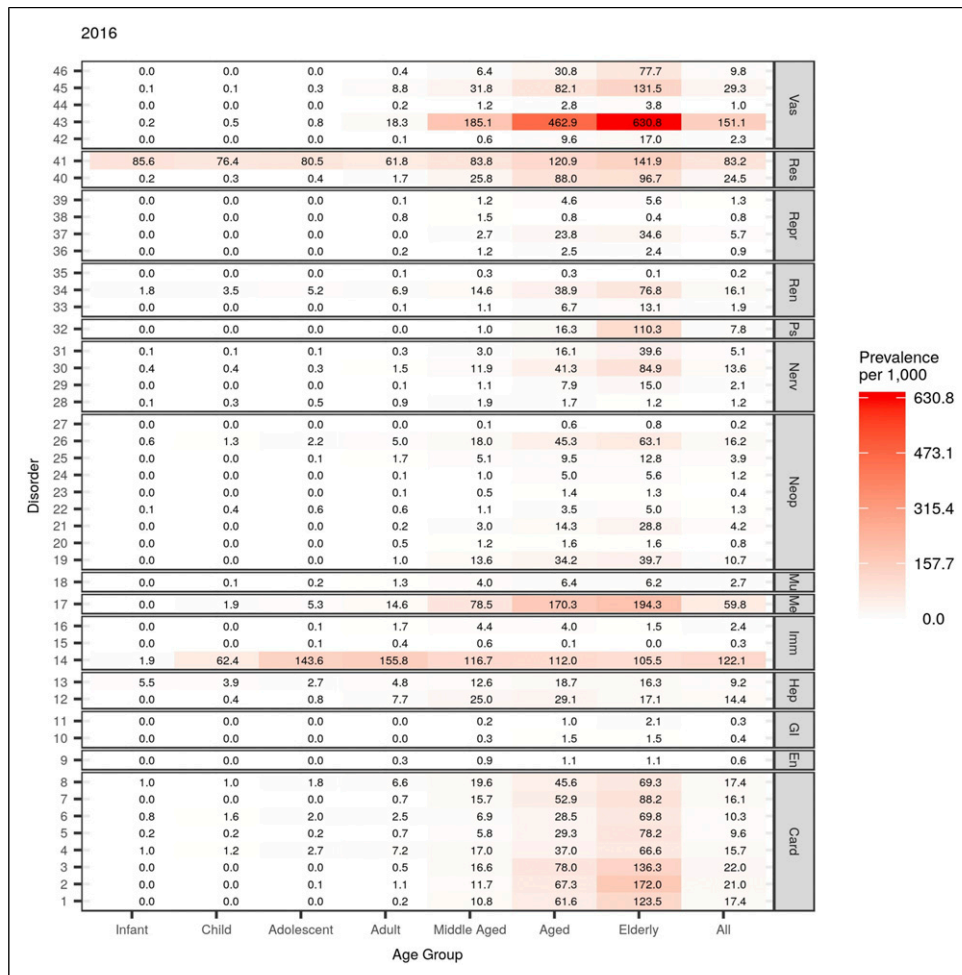
**Table 1.** Demographics of the dataset between 2005 and 2016.

Proportion (%)	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016
<b>Sex</b>												
Male	49.9	49.9	49.9	49.9	49.8	49.8	49.7	49.7	49.6	49.6	49.6	49.6
Female	50.1	50.1	50.1	50.1	50.2	50.2	50.3	50.3	50.4	50.4	50.4	50.4
<b>Country</b>												
England	70.1	69.8	68.9	68.7	67.9	67.2	66.7	65.9	63.3	59.7	51.1	45.1
Northern Ireland	4.3	4.4	4.6	4.4	4.5	4.6	4.7	4.7	5.1	5.5	6.8	7.6
Scotland	15.6	15.6	16.0	16.3	16.7	16.9	17.0	17.4	18.7	20.5	24.9	27.8
Wales	9.9	10.1	10.5	10.7	11.0	11.4	11.6	12.0	12.9	14.2	17.2	19.4
<b>Age Group</b>												
Infant	0.9	0.9	0.9	0.9	1.0	0.9	1.0	1.0	1.0	0.9	0.9	0.9
Child	12.5	12.5	12.5	12.5	12.6	12.6	12.7	12.8	12.8	13.0	13.0	13.1
Adolescent	6.6	6.6	6.6	6.6	6.6	6.6	6.7	6.7	6.7	6.8	6.8	6.8
Adult	35.6	35.4	35.2	34.9	34.7	34.3	34.0	33.8	33.3	33.0	32.9	33.1
Middle-Aged	27.2	27.4	27.6	27.7	27.7	27.9	27.8	27.6	27.6	27.6	27.7	27.7
Aged	12.5	12.4	12.5	12.6	12.7	12.8	13.1	13.4	13.7	13.8	13.9	13.7
Elderly	4.7	4.7	4.7	4.7	4.7	4.8	4.9	4.9	4.9	4.9	4.8	4.8
<b>Multimorbidity</b>												
0	63.6	62.9	62.4	62.0	61.7	61.4	61.1	60.9	60.5	60.3	60.2	60.4
1	21.8	22.0	22.1	22.3	22.3	22.4	22.4	22.4	22.5	22.5	22.5	22.4
2	7.9	8.1	8.3	8.4	8.5	8.6	8.7	8.8	8.9	8.9	8.9	8.9
3	3.3	3.5	3.6	3.6	3.7	3.8	3.8	3.9	4.0	4.0	4.1	4.0
4	1.7	1.8	1.8	1.9	1.9	1.9	2.0	2.0	2.0	2.1	2.1	2.1
5	0.9	0.9	0.9	1.0	1.0	1.0	1.0	1.0	1.1	1.1	1.1	1.1
6	0.4	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.6	0.6	0.6
7	0.2	0.2	0.2	0.2	0.2	0.3	0.3	0.3	0.3	0.3	0.3	0.3
8+	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.3	0.3	0.3
Total (millions)	4.84	4.93	5.04	5.11	5.07	5.01	4.97	4.92	4.62	4.23	3.51	3.15

high prevalence across all age groups, with the former occurring especially frequently in the Adolescent age group (13–19 years old). We observe the onset of metabolic and cardiovascular diseases in the Middle-Aged and older age groups, in particular, diabetes and hypertension. Not surprisingly, diseases such as dementia, kidney diseases, and stroke occur most frequently in the oldest patients (65 years and above). We observe an increasing trend in prevalence for some diseases. For example, the prevalence of diabetes in the Aged age group (65–80 years old) increased by almost 35% over the decade studied. In contrast, the prevalence of diseases such as angina fell over the study period. In [Table 2](#), we summarize the lift and odds ratio of the top ten most frequently co-occurring pairs of diseases in each age group in 2016. (See [Supplementary](#)

[Material C](#) for other years.) In all age groups, asthma occurs in combination with other respiratory-related diseases approximately two times more often than expected by chance (i.e., the lift is greater than 2.0). Additionally, the estimated odds ratios, both unadjusted and adjusted, indicate that patients with asthma are at least twice as likely to have other respiratory conditions at the same time, and vice versa.

Hypertension is most associated with a second condition in the older age groups, although most pairs do not necessarily occur more frequently than by chance. The combination of hypertension and diabetes stands out with a relatively high lift and an odds ratio that is greater than 2.0. Angina and coronary artery disease (CAD) also demonstrate a strong association in the Aged and Elderly age groups with unusually high lift and odds ratio.



**Figure 3.** Single-disease prevalence by age group in 2016. See Figures 1 and 2 for disorder to index mapping and MedDRA SOC abbreviations. See Supplementary Material C for breakout by age group and by year.

To better visualize the data, we plot the lift of all combinations of disease pairs in heat maps, stratified by MedDRA system organ classes. (See Figure 4 and Supplementary Material C for other age groups and years.) The co-occurrence of cardiac-cardiac and cardiac-respiratory disorders is a major risk across all age groups. We observe significant coupling between cardiac and hepatobiliary disorders in the Adolescent and Child (2–13 years old) age groups. On the other hand, combinations of cardiac-vascular and cardiac-metabolic disorders are the most dominant in the Middle-Aged and older age groups. We observe the same general patterns across time.

The proportion of patients with three or more co-occurring disorders is small in the younger age groups. For patients aged 45 years and older, triplets involving angina, CAD, hypertension, diabetes and MI occur most frequently with high lift, suggesting strong correlations between these diseases (see Table 3).

### Multimorbidity networks

In Figures 5 and 6, we plot the undirected and directed multimorbidity networks observed in the Aged age group in 2016. (See Supplementary Material C for other age groups and years.) Instead of a force-directed layout, we place the nodes in fixed positions around a circle to allow easy visualization of temporal changes in connections and clusters when comparing plots from different years. The edge thickness is proportional to the lift between each disease pair. Apart from single-node clusters, the communities detected using modularity maximization are given different colors.

In Tables 4 and 5, we identify clusters that remain relatively stable throughout the years in undirected and directed multimorbidity networks, respectively. We find between 1 and 4 clusters for each age group. The number of diseases in each cluster ranges between 2 and 12. In general, the communities found in Adolescent and younger patients

**Table 2.** Lift and odds ratio of the top 10 most prevalent multimorbidity pairs in 2016. See [Supplementary Material C](#) for other years. We include only the top four for the Infant age group due to the small sample size.

Age Group	Disease 1	Disease 2	N	Lift	Unadj OR (95% CI)	Adj OR (95%CI)
Infant	Liver-related	Other respiratory disease	26	1.9	2.1 (1.4, 3.2)	2.0 (1.3, 3.1)
Infant	Asthma	Other respiratory disease	15	3.3	4.2 (2.3, 7.7)	3.9 (2.2, 7.2)
Infant	Kidney disease	Other respiratory disease	10	2.3	2.6 (1.3, 5.2)	2.0 (1.0, 4.2)
Infant	Other cardiac	Other respiratory disease	6	2.5	2.9 (1.2, 7.2)	2.6 (0.9, 7.1)
Child	Asthma	Other respiratory disease	4,759	2.4	3.1 (3.0, 3.2)	3.1 (3.0, 3.2)
Child	Liver-related	Other respiratory disease	198	1.6	1.7 (1.5, 2.0)	1.6 (1.4, 1.9)
Child	Kidney disease	Other respiratory disease	159	1.5	1.5 (1.3, 1.8)	1.4 (1.1, 1.6)
Child	Asthma	Kidney disease	112	1.3	1.3 (1.1, 1.6)	1.0 (0.8, 1.2)
Child	Asthma	Liver-related	109	1.1	1.1 (0.9, 1.3)	1.1 (0.9, 1.4)
Child	Heart valve disorder	Other respiratory disease	104	2.1	2.4 (1.9, 2.9)	2.1 (1.7, 2.6)
Child	Cardiac arrhythmia	Other respiratory disease	75	2.0	2.1 (1.7, 2.7)	1.9 (1.4, 2.4)
Child	Other cardiac	Other respiratory disease	69	2.3	2.5 (1.9, 3.3)	2.2 (1.7, 2.9)
Child	Asthma	Diabetes	58	1.2	1.2 (0.9, 1.6)	0.9 (0.7, 1.2)
Child	Asthma	Heart valve disorder	51	1.3	1.3 (1.0, 1.7)	1.0 (0.8, 1.4)
Adolescent	Asthma	Other respiratory disease	5,272	2.1	3.0 (2.9, 3.1)	2.9 (2.8, 3.1)
Adolescent	Asthma	Kidney disease	187	1.2	1.2 (1.0, 1.4)	1.1 (0.9, 1.3)
Adolescent	Asthma	Diabetes	157	1.0	1.0 (0.8, 1.2)	0.9 (0.8, 1.1)
Adolescent	Kidney disease	Other respiratory disease	115	1.3	1.3 (1.1, 1.6)	1.2 (1.0, 1.4)
Adolescent	Asthma	Liver-related	112	1.4	1.5 (1.2, 1.8)	1.3 (1.1, 1.6)
Adolescent	Diabetes	Other respiratory disease	99	1.1	1.1 (0.9, 1.4)	1.1 (0.9, 1.4)
Adolescent	Asthma	Cardiac arrhythmia	96	1.2	1.2 (0.9, 1.5)	1.1 (0.9, 1.4)
Adolescent	Asthma	Heart valve disorder	90	1.5	1.6 (1.3, 2.0)	1.5 (1.2, 1.9)
Adolescent	Liver-related	Other respiratory disease	81	1.8	1.9 (1.5, 2.4)	1.7 (1.4, 2.2)
Adolescent	Asthma	Other cardiac	72	1.3	1.4 (1.1, 1.8)	1.2 (1.0, 1.6)
Adult	Asthma	Other respiratory disease	22,006	2.2	3.1 (3.1, 3.2)	3.0 (3.0, 3.1)
Adult	Asthma	Hypertension	3,296	1.1	1.1 (1.1, 1.2)	1.2 (1.2, 1.3)
Adult	Asthma	Diabetes	2,738	1.2	1.2 (1.1, 1.2)	1.2 (1.1, 1.2)
Adult	Diabetes	Hypertension	2,544	9.1	12.2 (11.7, 12.8)	7.4 (7.0, 7.7)
Adult	Asthma	PAD	1,768	1.2	1.3 (1.2, 1.4)	1.3 (1.2, 1.4)
Adult	Asthma	Cardiac arrhythmia	1,571	1.4	1.5 (1.4, 1.5)	1.4 (1.3, 1.5)
Adult	Hypertension	Other respiratory disease	1,549	1.3	1.3 (1.3, 1.4)	1.2 (1.2, 1.3)
Adult	Asthma	Liver disease	1,431	1.2	1.2 (1.1, 1.3)	1.2 (1.1, 1.3)
Adult	Asthma	Other cardiac	1,362	1.3	1.3 (1.3, 1.4)	1.3 (1.2, 1.3)
Adult	Diabetes	Other respiratory disease	1,323	1.4	1.5 (1.4, 1.5)	1.4 (1.3, 1.4)
Middle-Aged	Diabetes	Hypertension	33,471	2.6	5.1 (5.0, 5.1)	4.0 (3.9, 4.1)
Middle-Aged	Asthma	Hypertension	21,942	1.2	1.2 (1.2, 1.3)	1.2 (1.2, 1.2)
Middle-Aged	Asthma	Other respiratory disease	18,175	2.1	2.8 (2.8, 2.9)	2.5 (2.5, 2.6)
Middle-Aged	Hypertension	Other respiratory disease	17,128	1.3	1.4 (1.4, 1.4)	1.2 (1.1, 1.2)
Middle-Aged	Asthma	Diabetes	10,114	1.3	1.3 (1.3, 1.4)	1.2 (1.2, 1.3)
Middle-Aged	Diabetes	Other respiratory disease	8,461	1.5	1.6 (1.6, 1.7)	1.4 (1.3, 1.4)
Middle-Aged	Asthma	COPD	7,841	3.0	4.3 (4.2, 4.4)	3.8 (3.7, 3.9)
Middle-Aged	Hypertension	PAD	7,770	1.5	1.7 (1.7, 1.8)	1.2 (1.2, 1.2)
Middle-Aged	Hypertension	Liver disease	6,922	1.7	2.1 (2.0, 2.2)	1.4 (1.4, 1.5)
Middle-Aged	COPD	Hypertension	6,010	1.4	1.6 (1.6, 1.7)	1.0 (0.9, 1.0)
Aged	Diabetes	Hypertension	50,105	1.5	3.0 (2.9, 3.0)	2.7 (2.7, 2.8)
Aged	Hypertension	Other respiratory disease	26,587	1.1	1.2 (1.2, 1.3)	1.1 (1.1, 1.1)
Aged	Asthma	Hypertension	24,320	1.1	1.2 (1.2, 1.2)	1.1 (1.1, 1.2)
Aged	CAD	Hypertension	19,479	1.2	1.6 (1.6, 1.7)	1.2 (1.2, 1.2)
Aged	Hypertension	PAD	19,203	1.2	1.4 (1.4, 1.4)	1.2 (1.1, 1.2)
Aged	COPD	Hypertension	18,529	1.1	1.1 (1.1, 1.1)	0.9 (0.9, 1.0)

(continued)

Table 2. (continued)

Age Group	Disease 1	Disease 2	N	Lift	Unadj OR (95% CI)	Adj OR (95%CI)
Aged	Atrial fibrillation	Hypertension	17,644	1.3	1.9 (1.8, 1.9)	1.4 (1.4, 1.5)
Aged	Angina	Hypertension	16,234	1.3	1.9 (1.8, 1.9)	1.4 (1.4, 1.4)
Aged	Angina	CAD	14,867	7.2	25.9 (25.2, 26.6)	16.7 (16.2, 17.2)
Aged	Asthma	Other respiratory disease	12,245	2.1	2.9 (2.8, 3.0)	2.4 (2.4, 2.5)
Elderly	Diabetes	Hypertension	22,244	1.2	2.2 (2.1, 2.2)	2.2 (2.1, 2.2)
Elderly	Atrial fibrillation	Hypertension	18,073	1.1	1.5 (1.4, 1.5)	1.3 (1.3, 1.4)
Elderly	Hypertension	Other respiratory disease	13,975	1.0	1.1 (1.1, 1.2)	1.1 (1.0, 1.1)
Elderly	CAD	Hypertension	13,586	1.1	1.2 (1.2, 1.2)	1.1 (1.0, 1.1)
Elderly	Hypertension	PAD	13,283	1.1	1.2 (1.2, 1.3)	1.1 (1.1, 1.2)
Elderly	Angina	Hypertension	12,524	1.1	1.3 (1.2, 1.3)	1.1 (1.1, 1.2)
Elderly	Angina	CAD	10,359	4.1	15.3 (14.8, 15.9)	11.6 (11.2, 12.1)
Elderly	Asthma	Hypertension	10,327	1.0	1.1 (1.1, 1.2)	1.1 (1.0, 1.1)
Elderly	Dementia	Hypertension	9,992	1.0	0.9 (0.9, 0.9)	0.8 (0.8, 0.8)
Elderly	Hypertension	Stroke	9,039	1.1	1.5 (1.4, 1.5)	1.3 (1.3, 1.4)

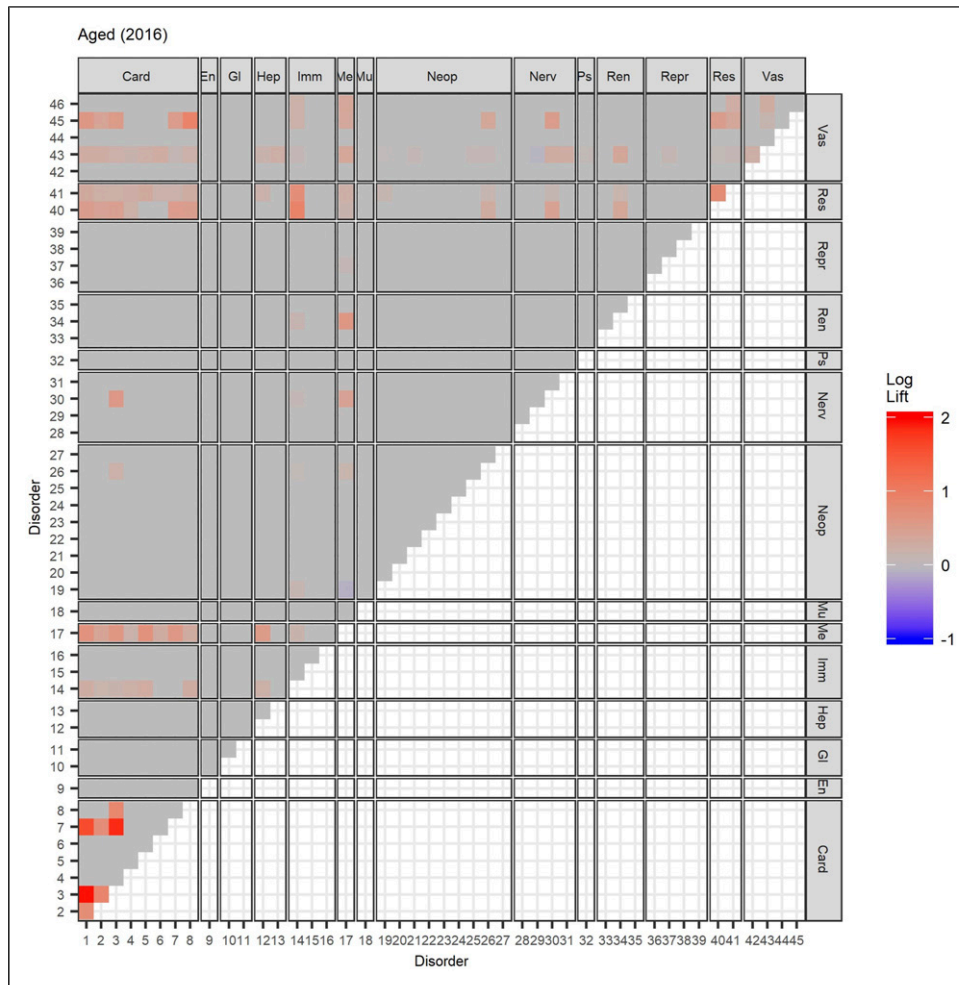


Figure 4. Heat map of lift of multimorbidity pairs in the Aged subgroup in 2016. See Figures 1 and 2 for disorder to index mapping and MedDRA SOC abbreviations. See Supplementary Material C for other age groups and years.



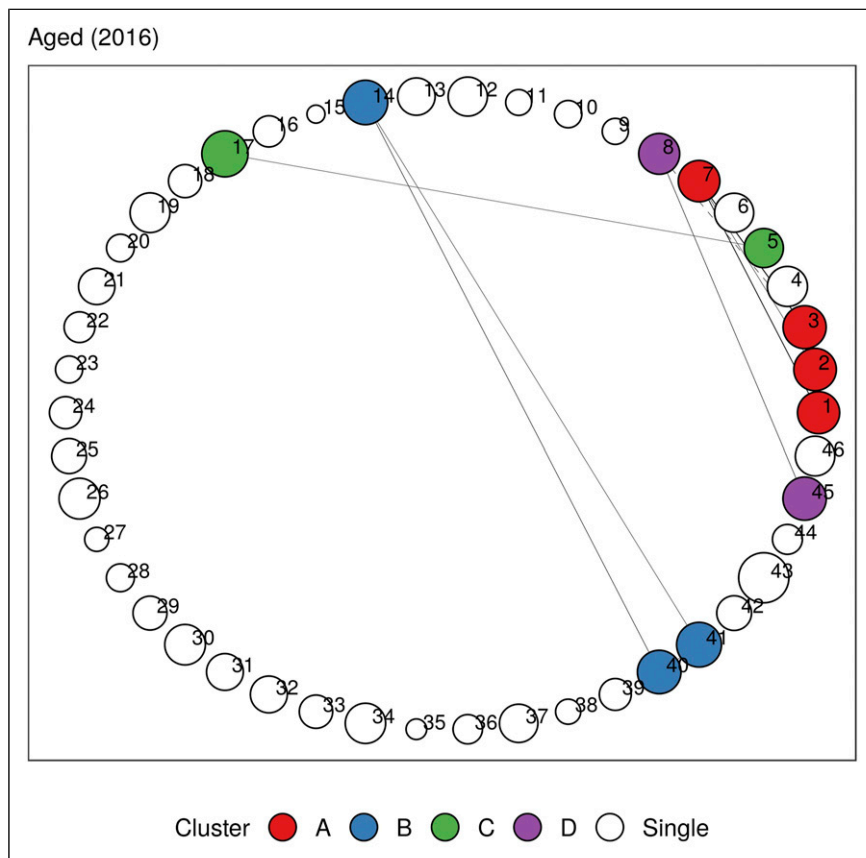
**Table 3.** Lift of the top 10 most prevalent multimorbidity triplets in 2016. See [Supplementary Material C](#) for other years. We exclude the Infant age group and include only the top five for the Child subgroup due to the small sample size.

Age Group	Disease 1	Disease 2	Disease 3	N	Lift
Child	Asthma	Kidney disease	Other respiratory disease	38	5.6
Child	Asthma	Liver-related	Other respiratory disease	35	4.6
Child	Asthma	Heart valve disorder	Other respiratory disease	20	6.6
Child	Asthma	Cardiac arrhythmia	Other respiratory disease	14	5.9
Child	Asthma	COPD	Other respiratory disease	13	23.9
Adolescent	Asthma	Kidney disease	Other respiratory disease	36	2.8
Adolescent	Asthma	Liver-related	Other respiratory disease	35	5.3
Adolescent	Asthma	Diabetes	Other respiratory disease	24	1.9
Adolescent	Asthma	Heart valve disorder	Other respiratory disease	23	4.7
Adolescent	Asthma	Cardiac arrhythmia	Other respiratory disease	20	3.0
Adolescent	Asthma	Other cancer	Other respiratory disease	13	2.4
Adolescent	Asthma	Other cardiac	Other respiratory disease	13	2.9
Adolescent	Asthma	COPD	Other respiratory disease	12	13.3
Adolescent	Asthma	Hypertension	Other respiratory disease	11	5.4
Adolescent	Asthma	Hypertension	Kidney disease	9	67.9
Adult	Asthma	Hypertension	Other respiratory disease	522	2.8
Adult	Asthma	Diabetes	Hypertension	477	11.0
Adult	Asthma	Diabetes	Other respiratory disease	462	3.2
Adult	Asthma	Other respiratory disease	PAD	329	3.7
Adult	Asthma	Cardiac arrhythmia	Other respiratory disease	253	3.5
Adult	Diabetes	Hypertension	Other respiratory disease	251	14.5
Adult	Asthma	Other cardiac	Other respiratory disease	238	3.6
Adult	Asthma	Liver disease	Other respiratory disease	228	3.0
Adult	Asthma	Kidney disease	Other respiratory disease	218	3.2
Adult	Asthma	Liver-related	Other respiratory disease	209	4.3
Middle-Aged	Asthma	Diabetes	Hypertension	5,094	3.4
Middle-Aged	Asthma	Hypertension	Other respiratory disease	4,621	2.9
Middle-Aged	Diabetes	Hypertension	Other respiratory disease	4,315	4.1
Middle-Aged	Asthma	Diabetes	Other respiratory disease	2,412	3.6
Middle-Aged	Diabetes	Hypertension	Liver Disease	2,394	7.6
Middle-Aged	Asthma	COPD	Other respiratory disease	2,342	10.7
Middle-Aged	Diabetes	Hypertension	PAD	2,323	5.8
Middle-Aged	CAD	Diabetes	Hypertension	2,282	10.9
Middle-Aged	Asthma	COPD	Hypertension	2,143	4.4
Middle-Aged	Angina	CAD	Hypertension	2,068	71.2
Aged	Angina	CAD	Hypertension	8,988	9.3
Aged	Diabetes	Hypertension	Other respiratory disease	7,738	1.9
Aged	CAD	Diabetes	Hypertension	7,417	2.8
Aged	Asthma	Diabetes	Hypertension	6,761	1.8
Aged	Asthma	Hypertension	Other respiratory disease	6,457	2.4
Aged	Angina	Diabetes	Hypertension	6,304	3.0
Aged	CAD	Hypertension	MI	6,171	7.5
Aged	Diabetes	Hypertension	PAD	5,975	2.1
Aged	Atrial fibrillation	Diabetes	Hypertension	5,417	2.4
Aged	Asthma	COPD	Hypertension	5,290	2.7
Elderly	Angina	CAD	Hypertension	6,910	4.3
Elderly	Atrial fibrillation	Diabetes	Hypertension	4,717	1.5
Elderly	CAD	Diabetes	Hypertension	4,301	1.7
Elderly	CAD	Hypertension	MI	4,237	3.7
Elderly	Atrial fibrillation	Heart failure	Hypertension	4,097	3.2

(continued)

**Table 3.** (continued)

Age Group	Disease 1	Disease 2	Disease 3	N	Lift
Elderly	Atrial fibrillation	CAD	Hypertension	3,951	1.8
Elderly	Angina	Diabetes	Hypertension	3,899	1.7
Elderly	Diabetes	Hypertension	Other respiratory disease	3,788	1.5
Elderly	Diabetes	Hypertension	PAD	3,561	1.5
Elderly	Angina	Atrial Fibrillation	Hypertension	3,333	1.7



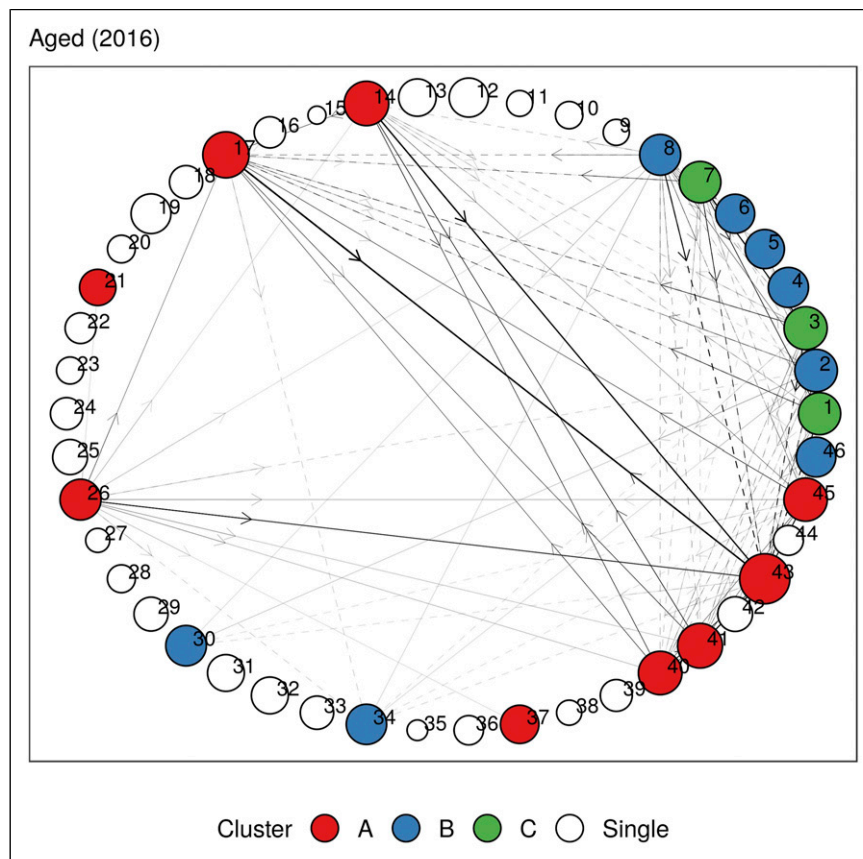
**Figure 5.** Undirected multimorbidity network in the Aged subgroup in 2016. Edge thickness is proportional to the lift between each disease pair. Intra-group edges and inter-group edges are represented by solid lines and dashed lines, respectively. Only communities with more than one node are colored. See [Figures 1](#) and [2](#) for mapping of disorder to index and MedDRA SOC abbreviations. See [Supplementary Material C](#) for other age groups and years.

can vary greatly from year to year compared to older age groups, where the clusters evolve very little over time. This is expected, given that only a small proportion of the former cohort has more than two co-occurring disorders, so the results are sensitive to small changes in prevalence each year.

A respiratory cluster of asthma, chronic obstructive pulmonary disease (COPD), and respiratory-related diseases appears to be present in all age groups in both undirected and directed graphs. Similarly, a vascular-metabolic-hepatobiliary-renal cluster that is characterized

by hypertension, diabetes, liver diseases, and kidney diseases, with the occasional appearance of cardiac disorders, is also present in almost all cohorts. As observed in previous analyses, we also find several clusters dominated by cardiovascular disorders such as angina, CAD, myocardial infarction (MI), atrial fibrillation, cardiac arrhythmia, heart failure, heart valve disorder (HVD), stroke, peripheral artery disease (PAD), and transient ischemic attack (TIA).

In [Tables 6](#) and [7](#), we summarize the top five diseases for each centrality measure. (See [Supplementary Material C](#) for



**Figure 6.** Directed multimorbidity network in the Aged subgroup in 2016. Edge thickness is proportional to the lift between each disease pair. Intra-group edges and inter-group edges are represented by solid lines and dashed lines, respectively. Only communities with more than one node are colored. See [Figures 1 and 2](#) for mapping of disorder to index and MedDRA SOC abbreviations. See [Supplementary Material C](#) for other age groups and years.

the full set of results.) The degree centrality and eigencentality for hypertension, diabetes, CAD, and angina are the highest when all age groups are aggregated in undirected multimorbidity networks. In the Adolescent and younger age groups, kidney disease shows both high degree centrality and eigencentality. Other important nodes include respiratory-related diseases and HVD, which have high degree centrality and high eigencentality, respectively. For the Adult and Middle-Aged age groups, hypertension and diabetes are the most central nodes with respect to both measures. In the Aged and Elderly age groups, we find that cardiac disorders make up all of the top five most connected nodes. Diseases with high out-degree centrality often lead to a second disease, while diseases with high in-degree centrality are often diagnosed following an earlier condition.

We observe similar results in directed networks. In general, hypertension, diabetes, and respiratory-related diseases demonstrate high in-degree centrality and eigencentality, while cardiac disorders show high out-degree centrality. In the Middle-Aged and younger age groups, asthma emerges as a new central node with high in-degree

centrality, while the top five diseases for the Aged and Elderly age groups remain dominated by cardiovascular diseases.

### Survival analysis

We summarize the dataset used for survival analysis in [Supplementary Material D](#). The sample consists of approximately 390,000 patients in the Aged and Elderly age groups for each year between 2010 and 2012. More than 50% of the patients are multimorbid. In terms of predicting five-year overall survival, we find the performance of the linear and nonlinear survival models explored to be very similar. We focus on the Cox model here due to its ease of interpretability. The model achieves a promising C-index of 0.81 (95% CI 0.80–0.81) on out-of-sample data in 2012. In addition, its calibration curves lay close to the ideal diagonal, indicating that the model is well calibrated, that is, the model does not systematically overestimate or underestimate survival rates in any of the quintiles. (See [Supplementary Material C](#) for plots.)

**Table 4.** Clusters identified in undirected multimorbidity networks in different age groups.

Age Group	Cluster 1	Cluster 2	Cluster 3	Cluster 4
Infant	Asthma, COPD, respiratory-related diseases	Cardiac arrhythmia, heart failure, HVD, cardiac-related diseases, kidney disease, hypertension		
Child	Cardiac arrhythmia, HVD, cardiac-related diseases, hypertension	Liver disease, liver-related diseases, encephalitis, stroke, kidney disease, hypertension, PAD		
Adolescent	Asthma, COPD, respiratory-related diseases	Cardiac arrhythmia, HVD, cardiac-related diseases	Liver disease, liver-related diseases, diabetes, leukemias, kidney disease, hypertension, PAD	
Adult	Asthma, COPD, respiratory-related diseases	Cardiac arrhythmia, cardiac-related diseases, kidney disease, PAD	HVD, liver disease, liver-related diseases, diabetes, lupus, hypertension	
Middle-Aged	CAD, MI, cardiac-related diseases, asthma, COPD, respiratory-related diseases, PAD	Angina, atrial fibrillation, heart failure, liver disease, liver-related diseases, diabetes, stroke, stroke-related diseases, TIA		
Aged	Cardiac-related diseases, PAD	Heart failure, diabetes	Asthma, COPD, respiratory-related diseases	Angina, atrial fibrillation, CAD, cardiac arrhythmia, MI, TIA
Elderly	Asthma, COPD, respiratory-related diseases	Angina, CAD, MI	Atrial fibrillation, cardiac arrhythmia, heart failure, HVD	

**Table 5.** Clusters identified in directed multimorbidity networks in different age groups.

Age Group	Cluster 1	Cluster 2	Cluster 3	Cluster 4
Infant	Cardiac arrhythmia, cardiac-related diseases, liver disease, liver-related diseases, respiratory-related diseases			
Child	Kidney disease, hypertension, PAD	Cardiac arrhythmia, heart failure, HVD, cardiac-related diseases	Liver-related diseases, asthma, diabetes, respiratory-related diseases	
Adolescent	Liver-related diseases, asthma, respiratory-related diseases	Cardiac arrhythmia, heart failure, HVD, cardiac-related diseases	Diabetes, kidney disease, hypertension, PAD	
Adult	Asthma, respiratory-related diseases	Liver disease, liver-related diseases	Diabetes, lupus, kidney disease, kidney-related diseases, hypertension	Atrial fibrillation, heart failure, HVD, cardiac-related diseases, lupus, stroke, PAD
Middle-Aged	Atrial fibrillation, cardiac arrhythmia, HVD, cardiac-related diseases, stroke	Angina, CAD, heart failure, MI, stroke, TIA		
Aged	Angina, CAD, MI	Atrial fibrillation, cardiac arrhythmia, heart failure, HVD, cardiac-related diseases, stroke, TIA	Liver disease, liver-related diseases, asthma, diabetes, breast cancer, colorectal cancer, cancer-related diseases, kidney disease, COPD, respiratory-related diseases, hypertension, PAD	
Elderly	Angina, CAD, MI, cardiac-related diseases	Asthma, diabetes, COPD, respiratory-related diseases, hypertension, PAD	Atrial fibrillation, cardiac arrhythmia, heart failure, HVD, stroke, stroke-related diseases, dementia, TIA	

**Table 6.** Centrality measures for top five diseases in undirected multimorbidity networks with mean computed over time.

All		Infant		Child		Adolescent	
Disease	Mean	Disease	Mean	Disease	Mean	Disease	Mean
<b>Degree Centrality</b>							
Hypertension	23.5	Respiratory-related	7.7	Kidney Disease	11.7	Kidney Disease	10.1
Diabetes	17.5	Cardiac-related	5.1	Liver-related	10.1	Diabetes	7.4
PAD	11.9	HVD	4.8	Respiratory-related	9.8	Respiratory-related	7.0
CAD	7.8	Liver-related	4.4	HVD	7.9	Liver-related	6.8
Angina	7.7	Cardiac Arrhythmia	3.7	Cardiac-related	6.5	Asthma	6.8
<b>Eigencentality</b>							
CAD	0.9	Cardiac-related	0.9	HVD	1.0	Kidney Disease	1.0
Diabetes	0.9	HVD	0.8	Kidney Disease	0.9	HVD	0.9
Hypertension	0.9	Hypertension	0.6	Cardiac-related	0.8	Cardiac-related	0.9
Angina	0.9	Cardiac Arrhythmia	0.6	Cardiac Arrhythmia	0.8	Liver Disease	0.8
PAD	0.8	Kidney Disease	0.5	Hypertension	0.8	Cardiac Arrhythmia	0.7
<b>Transitivity</b>							
	0.32		0.29		0.37		0.40
Adult		Middle-Aged		Aged		Elderly	
Disease	Mean	Disease	Mean	Disease	Mean	Disease	Mean
<b>Degree Centrality</b>							
Hypertension	12.7	Diabetes	10.8	CAD	5.4	Heart Failure	3.5
PAD	8.5	Hypertension	9.5	Angina	4.7	CAD	3.1
Cardiac-related	7.8	PAD	6.8	Atrial Fibrillation	3.0	Angina	3.0
Kidney Disease	7.6	COPD	5.9	MI	3.0	Atrial Fibrillation	2.5
Liver Disease	6.7	CAD	5.7	Cardiac-related	2.8	MI	2.3
<b>Eigencentality</b>							
Hypertension	1.0	Diabetes	0.8	CAD	1.0	CAD	1.0
Cardiac-related	0.8	CAD	0.8	Angina	0.9	Angina	0.9
Liver Disease	0.8	Angina	0.7	MI	0.8	MI	0.8
Kidney Disease	0.8	Hypertension	0.6	Atrial Fibrillation	0.5	Heart Failure	0.8
Liver-related	0.8	PAD	0.6	Cardiac-related	0.4	Atrial Fibrillation	0.4
<b>Transitivity</b>							
	0.66		0.31		0.59		0.57

We extract the top ten coefficients in the Cox model to identify specific risk factors (see [Supplementary Material D](#)). To correct for multiple testing, we perform the Benjamini–Hochberg adjustment with a 5% false discovery rate for identifying significant factors. Apart from cancers, we find the presence of multimorbidity to be a strong adverse risk factor, that is, the higher the number of co-occurring chronic conditions, the greater the mortality risk. For example, the hazard ratio of having four or more

chronic conditions is 2.44 (95% CI 2.22–2.69). We also find a high IMD, corresponding to a lower socioeconomic status, to be significantly associated with increased risk, although this factor is not in the top ten coefficients.

## Discussion

With multimorbidity becoming the norm rather than the exception,<sup>2,12,17,25,49,50</sup> the management of multiple chronic

**Table 7.** Centrality measures for top five diseases in directed multimorbidity networks with mean computed over time. We exclude eigencentralities that are close to zero.

All		Infant		Child		Adolescent	
Disease	Mean	Disease	Mean	Disease	Mean	Disease	Mean
<b>In-degree Centrality</b>							
Diabetes	11.0	Respiratory-related	5.3	Asthma	10.9	Asthma	11.1
Respiratory-related	11.0	Cardiac-related	0.5	Respiratory-related	10.7	Respiratory-related	10.9
Hypertension	11.0	Kidney Disease	0.4	Kidney Disease	3.4	Hypertension	3.9
COPD	10.8	HVD	0.3	HVD	3.1	Kidney Disease	3.1
PAD	10.1	Liver-related	0.3	Cardiac Arrhythmia	2.9	HVD	2.7
<b>Out-degree Centrality</b>							
CAD	15.3	Heart Failure	1.3	Cardiac-related	6.3	Hypertension	7.5
Atrial Fibrillation	15.2	Hypertension	1.3	Hypertension	5.7	Liver Disease	6.5
Angina	14.9	Liver-related	1.0	Liver Disease	5.3	HVD	6.0
Cardiac-related	14.0	Stroke	0.9	Leukemias	5.2	Cardiac-related	5.8
Hypertension	12.9	Cardiac Arrhythmia	0.7	HVD	4.4	Cancer-related	3.8
<b>Eigencentality</b>							
Hypertension	1.0			Asthma	0.9	Asthma	0.9
Diabetes	0.4			Respiratory-related	0.4	Respiratory-related	0.4
CAD	0.4						
Respiratory-related	0.4						
Angina	0.3						
<b>Transitivity</b>							
	0.76		0.14		0.57		0.57
<b>Adult</b>		<b>Middle-Aged</b>		<b>Aged</b>		<b>Elderly</b>	
Disease	Mean	Disease	Mean	Disease	Mean	Disease	Mean
<b>In-degree Centrality</b>							
Asthma	11.0	Asthma	11.0	Atrial Fibrillation	11.0	Atrial Fibrillation	11.0
Diabetes	11.0	Diabetes	11.0	Diabetes	11.0	Hypertension	11.0
Respiratory-related	11.0	Respiratory-related	11.0	Respiratory-related	11.0	Diabetes	10.6
Hypertension	11.0	Hypertension	11.0	Hypertension	11.0	PAD	10.4
PAD	11.0	PAD	11.0	PAD	11.0	CAD	10.3
<b>Out-degree Centrality</b>							
Cardiac-related	13.4	Cardiac-related	16.8	Cardiac-related	14.2	CAD	13.1
Kidney Disease	12.3	CAD	14.6	CAD	13.8	Hypertension	12.5
HVD	12.3	Hypertension	14.5	Atrial Fibrillation	13.1	Angina	12.4
Cancer-related	10.4	PAD	10.9	Angina	12.0	PAD	12.4
Hypertension	10.3	MI	10.7	Hypertension	11.7	Atrial Fibrillation	12.3
<b>Eigencentality</b>							
Asthma	1.0	Hypertension	1.0	Hypertension	1.0	Hypertension	1.0
Respiratory-related	0.7	Diabetes	0.5	Diabetes	0.4	CAD	0.4
Hypertension	0.5	Respiratory-related	0.4	CAD	0.4	Angina	0.3

(continued)

**Table 7.** (continued)

Adult		Middle-Aged		Aged		Elderly	
Disease	Mean	Disease	Mean	Disease	Mean	Disease	Mean
Diabetes	0.3	Asthma	0.3	Angina	0.3	Atrial Fibrillation	0.3
PAD	0.2	CAD	0.3	Respiratory-related	0.3	Diabetes	0.3
	0.82		0.70		0.78		0.75

diseases in older adults is a major challenge facing healthcare systems worldwide. It is clear that a better understanding of the epidemiology of multimorbidity is required to develop more effective preventive interventions and better primary medical care for multimorbid patients. In this paper, we use data-driven methods to characterize multimorbidity patterns in different demographic groups and their evolution over the past decade, using a large, representative electronic medical records database consisting of over 4.5 million patients.

Consistent with other studies, we find that the prevalence and severity of multimorbidity increase substantially with age. In addition, we observe social inequalities in multimorbidity, with patients in socioeconomically deprived areas more likely to be multimorbid.<sup>11,12,38,49,51–53</sup> Our findings also support the role of hypertension as an important risk factor in older adults, as reported in the literature.<sup>2,40,54,55</sup> Hypertension is one of the most prevalent and most central chronic conditions in our dataset, and one that serves as an important bridge between many diseases in our networks. Other trends identified in our analysis, such as the falling prevalence of angina<sup>56–59</sup> and the growing prevalence of diabetes,<sup>60</sup> are also well documented in previously published population studies.

In our pairwise analysis, we find strong association between multiple pairs of chronic conditions, including between asthma and respiratory-related diseases<sup>61,62</sup> in the Adolescent age group, between hypertension and diabetes<sup>28,63–65</sup> and between CAD and angina<sup>66</sup> among older patients, and between cardiovascular and respiratory disorders in all age groups.<sup>2</sup> Triplets involving cardiovascular and metabolic disorders, such as CAD, hypertension, and diabetes, also occurred more frequently than expected by chance.<sup>2,25,28,67,68</sup>

Our network analysis further identified several meaningful communities that are common across all demographics, including a respiratory cluster (e.g., asthma and COPD),<sup>69</sup> a cardiovascular cluster,<sup>19,70,71</sup> and a mixed cardiovascular-renal-metabolic cluster,<sup>39,72–74</sup> all of which are supported by either established pathophysiological mechanisms or shared risk factors. For example, it is well known that cardiovascular diseases are one of the most common complications of diabetes. While we do not find

any particular multimorbidity pattern to have a significant effect on mortality, our models do indeed verify the substantial burden of multimorbidity (as quantified by the number of co-occurring chronic conditions) on overall survival in older patients.<sup>7,12,26,75,76</sup>

However, we must emphasize that our results do not necessarily imply any causal link between diseases identified to be in the same cluster. The association might be attributable to shared risk factors (e.g., smoking) or other adverse events, and any temporal relationships to be inferred from the multimorbidity directed networks might be administrative in nature (e.g., incomplete medical records that are rectified in subsequent visits) or biased by delayed diagnosis.

In general, the lack of an accepted standard for defining multimorbidity makes it difficult for any meaningful comparison of results across different studies.<sup>77,78</sup> Moreover, because results can be highly dependent on the study population, the disease ontology used, and the number of chronic conditions considered, it is not uncommon for studies to report seemingly conflicting findings. In this paper, we consider a wide range of demographic groups and a total of 46 morbidities, which is more than most similar studies,<sup>11</sup> and well above the minimum of 11–12 as recommended by systematic reviews in this field of research.<sup>78,79</sup> In addition, our findings are largely consistent with existing studies in the medical literature.

Lastly, we note that cancer appears to be under represented in the THIN database. This is because many cancer patients are treated separately in cancer centers under the care of specialized clinical teams. Unfortunately, data on such patients rarely make their way back to the primary care clinics where the THIN data is collected, leading to a gap in this area.

## Conclusions

Current healthcare systems are largely centered on single-disease approaches to treatment, resulting in the fragmentation of care and a lack of continuity in the management of multiple diseases. Even most clinical trials exclude multimorbid patients. Because multimorbidity is more common



in disadvantaged groups, the current structure exacerbates health inequalities in society.

In this paper, we apply statistical methods and network analysis to characterize multimorbidity associations in the general UK population using a large electronic medical records database spanning the years 2005–2016. We find that the proportion of the population with multimorbidity has increased over the last decade, and the prevalence and severity of multimorbidity increase substantially with age. We identify strongly associated comorbid pairs of cardiovascular and cardiac-metabolic disorders. In addition, our clustering algorithm reveals three principal clusters: a respiratory cluster, a cardiovascular cluster, and a mixed cardiovascular-renal-metabolic cluster. In our directed network analysis, hypertension, diabetes, and respiratory-related diseases demonstrate high in-degree centrality, while cardiac disorders show high out-degree centrality. Our findings largely confirm and expand on the results of existing studies in the literature. We believe that our results contribute to a better understanding of multimorbidity that may be useful for the early detection and prevention of comorbidities, for example, prescribing lifestyle interventions (i.e., adopting healthy dietary and exercise regimens) to hypertension patients as a preventive measure for diabetes.<sup>80</sup>

There is a pressing need for a universal framework that standardizes the way that multimorbidity is assessed (e.g., the appropriate number of diseases and the choice of chronic conditions to include) in order to facilitate comparisons between studies and populations. With the “Omics” revolution, the combination of phenotypic, genomic, and epigenomic data has the potential to provide deeper insights into the underlying pathophysiological associations between comorbid diseases. Unfortunately, the availability of such linked datasets remains very limited. Further research is also needed to better understand the impact of multimorbidity on different health outcomes, such as quality of life and healthcare costs, in order to align the healthcare system more closely to the needs to multimorbid patients.

### Acknowledgments

We thank Christoph Nabholz for supporting the project and Jayna Cummings for editorial support. Research support from the MIT Laboratory for Financial Engineering is gratefully acknowledged. The views and opinions expressed in this article are those of the authors only, and do not necessarily represent the views and opinions of any institution or agency, any of their affiliates or employees, or any of the individuals acknowledged above.

### Author contributions

Conceptualization, K.W.S., C.H.W., J.G., and A.W.L.; resources, J.G., and A.W.L.; methodology, K.W.S., C.H.W., J.G., and A.W.L.; software, K.W.S.; formal analysis, K.W.S.; writing—original draft, K.W.S. and A.W.L.; writing—review and editing,

K.W.S., C.H.W., J.G., and A.W.L.; supervision, A.W.L.; project administration, J.G. and A.W.L.

### Declaration of conflicting interests

The author(s) declared the following potential conflicts of interest with respect to the research, authorship, and/or publication of this article: K.W.S. and C.H.W. declare no competing interests. J.G. is an employee of Swiss Re and declares no competing interests. A.W.L. reports personal investments in private biotech companies, biotech venture capital funds, and mutual funds. A.W.L. is a co-founder and partner of QLS Advisors, a healthcare analytics and consulting company; an advisor to Apricity Health, Aracari Bio, BrightEdge Ventures, Enable Medicine, FINRA, Lazard, NIH/NCATS, Quantile Health, SalioGen Therapeutics, Swiss Finance Institute, and Thalès; and a director of AbCellera, Annual Reviews, Atomwise, BridgeBio Pharma, and Roivant Sciences. During the most recent six-year period, A.W.L. has received speaking/consulting fees, honoraria, or other forms of compensation from: AbCellera, AlphaSimplex Group, Annual Reviews, Apricity Health, Aracari Bio, Atomwise, Bernstein Fabozzi Jacobs Levy Award, BridgeBio Pharma, Cambridge Associates, Chicago Mercantile Exchange, Enable Medicine, Financial Times, Harvard University, IMF, Journal of Investment Management, Lazard, National Bank of Belgium, New Frontier Advisors/Markowitz Award, Oppenheimer, Princeton University Press, Q Group, QLS Advisors, Quantile Health, Research Affiliates, Roivant Sciences, SalioGen, Swiss Finance Institute, and WW Norton.

### Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: No direct funding was received for this study; general research support was provided by the MIT Laboratory for Financial Engineering and its sponsors. The authors were personally salaried by their institutions during the period of writing (though no specific salary was set aside or given for the writing of this paper).

### Data availability

The data that support the findings of this study are available from The Health Improvement Network (THIN; <https://www.the-health-improvement-network.com/en/>) but restrictions apply to the availability of these data, which were used under license for the current study, and so are not publicly available. Data are however available from the authors upon reasonable request and with permission of THIN.

### ORCID iDs

Kien Wei Siah  <https://orcid.org/0000-0002-7837-7030>

Chi Heem Wong  <https://orcid.org/0000-0002-4899-5022>

Andrew W Lo  <https://orcid.org/0000-0003-2944-7773>

## Supplemental material

Supplementary material for this article is available online.

## References

- World Health Organization. *Multimorbidity Technical Series on Safer Primary Care Multimorbidity: Technical Series on Safer Primary Care*, 2016.
- Schäfer I, Kaduszkiewicz H, Wagner HO, et al. Reducing complexity: A visualisation of multimorbidity by combining disease clusters and triads. *BMC Public Health* 2014; 14: 1285.
- Kadam U and Croft P. Clinical multimorbidity and physical function in older adults: a record and health status linkage study in general practice. *Fam Pract* 2007; 24: 412–419.
- Laux G, Kuehlein T, Rosemann T, et al. Co- and multimorbidity patterns in primary care based on episodes of care: Results from the German CONTENT project. *BMC Health Serv Res* 2008; 8: 14.
- Fung CH, Setodji CM, Kung FY, et al. The relationship between multimorbidity and patients' ratings of communication. *J Gen Intern Med* 2008; 23: 788–793.
- Schoenberg NE, Kim H, Edwards W, et al. Burden of common multiple-morbidity constellations on out-of-pocket medical expenditures among older adults. *Gerontologist* 2007; 47: 423–437.
- Gijsen R, Hoeymans N, Schellevis FG, et al. Causes and consequences of comorbidity: a review. *J Clin Epidemiol* 2001; 54: 661–674.
- Salisbury C, Johnson L, Purdy S, et al. Epidemiology and impact of multimorbidity in primary care: a retrospective cohort study. *Br J Gen Pract* 2011; 61: e12–e21.
- Wolff JL, Starfield B and Anderson G. Prevalence, expenditures, and complications of multiple chronic conditions in the elderly. *Arch Intern Med* 2002; 162: 2269.
- Fortin M, Lapointe L, Hudon C, et al. Multimorbidity and quality of life in primary care: a systematic review. *Health Qual Life Outcomes* 2004; 2: 1–12.
- Barnett K, Mercer SW, Norbury M, et al. Epidemiology of multimorbidity and implications for health care, research, and medical education: a cross-sectional study. *Lancet* 2012; 380: 37–43.
- Marengoni A, Angleman S, Melis R, et al. Aging with multimorbidity: a systematic review of the literature. *Ageing Res Rev* 2011; 10: 430–439.
- Loza E, Jover JA, Rodriguez L, et al. Multimorbidity: prevalence, effect on quality of life and daily functioning, and variation of this effect when one condition is a rheumatic disease. *Semin Arthritis Rheum* 2009; 38: 312–319.
- Crentsil V, Ricks MO, Xue QL, et al. A pharmacoepidemiologic study of community-dwelling, disabled older women: Factors associated with medication use. *Am J Geriatr Pharmacother* 2010; 8: 215–224.
- Walker AE. Multiple chronic diseases and quality of life: patterns emerging from a large national sample, Australia. *Chronic Illn* 2007; 3: 202–218.
- Van den Akker M, Buntix F, Metsemakers JFM, et al. Multimorbidity in general practice: prevalence, incidence, and determinants of co-occurring chronic and recurrent diseases. *J Clin Epidemiol* 1998; 51: 367–375.
- Fortin M. Prevalence of multimorbidity among adults seen in family practice. *Ann Fam Med* 2005; 3: 223–228.
- Boyd CM, Darer J, Boult C, et al. Clinical practice guidelines and quality of care for older patients with multiple comorbid diseases. *JAMA* 2005; 294: 716.
- Marengoni A, Rizzuto D, Wang H-X, et al. Patterns of Chronic Multimorbidity in the Elderly Population. *J Am Geriatr Soc* 2009; 57: 225–230.
- Marengoni A, Bonometti F, Nobili A, et al. In-hospital death and adverse clinical events in elderly patients according to disease clustering: The REPOSI study. *Rejuvenation Res* 2010; 13: 469–477.
- Redelmeier DA, Tan SH and Booth GL. The Treatment of Unrelated Disorders in Patients with Chronic Medical Diseases. *N Engl J Med* 1998; 338: 1516–1520.
- Ioakeim-Skoufa I, Poblador-Plou B, Carmona-Pirez J, et al. Multimorbidity Patterns in the General Population: Results from the EpiChron Cohort Study. *Int J Environ Res Public Heal* 2020; 17: 4242.
- Guisado-Clavero M, Roso-Llorach A, Lopez-Jimenez T, et al. Multimorbidity patterns in the elderly: A prospective cohort study with cluster analysis. *BMC Geriatr* 2018; 18: 1–11.
- Nicholson K, Bauer M, Terry A, et al. The multimorbidity cluster analysis tool: identifying combinations and permutations of multiple chronic diseases using a record-level computational analysis. *BMJ Heal Care Informatics* 2017; 24: 339–343.
- Schäfer I, von Leitner E-C, Schon G, et al. Multimorbidity patterns in the elderly: a new approach of disease clustering identifies complex interrelations between chronic conditions. *PLoS One* 2010; 5: e15941.
- Ferrer A, Formiga F, Sanz H, et al. Multimorbidity as specific disease combinations, an important predictor factor for mortality in octogenarians: the Octabaix study. *Clin Interv Aging* 2017; 12: 223–231.
- Diederichs C, Berger K and Bartels DB. The measurement of multiple chronic diseases--a systematic review on existing multimorbidity indices. *Journals Gerontol Ser A Biol Sci Med Sci* 2011; 66A: 301–311.
- Kirchberger I, Meisinger C, Heier M, et al. Patterns of multimorbidity in the aged population. Results from the KORA-age study. *PLoS One* 2012; 7: e30556.
- Bisquera A, Gulliford M, Dodhia H, et al. Identifying longitudinal clusters of multimorbidity in an urban setting: a population-based cross-sectional study. *Lancet Reg Heal - Eur* 2021; 3: 100047.

30. Violán C, Foguet-Boreu Q, Hermosilla-Perez E, et al. Comparison of the information provided by electronic health records data and a population health survey to estimate prevalence of selected health conditions and multimorbidity. *BMC Public Health* 2013; 13: 1–10.
31. Hassaine A, Canoy D, Solares JRA, et al. Learning multimorbidity patterns from electronic health records using non-negative matrix factorisation. *J Biomed Inform* 2020; 112: 103606.
32. Cezard G, McHale CT, Sullivan F, et al. Studying trajectories of multimorbidity: a systematic scoping review of longitudinal approaches and evidence. *BMJ Open* 2021; 11: e048485.
33. The Health Improvement Network. *The health improvement network*, <https://www.the-health-improvement-network.com/en/>
34. Kastner M, Wilczynski NL, Walker-Dilks C, et al. Age-specific search strategies for medline. *J Med Internet Res* 2006; 8.
35. National Health Services. *Read Codes*, 2020, <https://digital.nhs.uk/services/terminology-and-classifications/read-codes>
36. Timmreck TC, Cole GE, James G and Butterworth DD. Health education and health promotion: a look at the jungle of supportive fields, philosophies and theoretical foundations. *Health Educ* 1987; 18: 23–28.
37. Ministry of Housing Communities, Local Government. *English Indices of Deprivation*, 2012, <https://www.gov.uk/government/collections/english-indices-of-deprivation>
38. Orueta JF, García-Álvarez A, García-Goñi M, et al. Prevalence and costs of multimorbidity by deprivation levels in the basque country: a population based study using health administrative databases. *PLoS One* 2014; 9: e89787.
39. Aguado A, Moratalla-Navarro F, López-Simarro F, et al. MorbiNet: multimorbidity networks in adult general population. Analysis of type 2 diabetes mellitus comorbidity. *Sci Rep* 2020; 10: 2416.
40. Feldman K, Stiglic G, Dasgupta D, et al. Insights into population health management through disease diagnoses networks. *Sci Rep* 2016; 6: 30465.
41. Leva F and Bitonti D. Network analysis of comorbidity patterns in heart failure patients using administrative data. *Epidemiol Biostat Public Heal* 2018; 15.
42. Liu J, Ma J, Wang J, et al. Comorbidity analysis according to sex and age in hypertension patients in China. *Int J Med Sci* 2016; 13: 99–107.
43. Brandes U, Delling D, Geartler M, et al. On modularity clustering. *IEEE Trans Knowl Data Eng* 2008; 20: 172–188.
44. Girvan M and Newman MEJ. Community structure in social and biological networks. *Proc Natl Acad Sci* 2002; 99: 7821–7826.
45. Newman MEJ. Fast algorithm for detecting community structure in networks. *Phys Rev E - Stat Physics, Plasmas, Fluids, Relat Interdiscip Top* 2004; 69: 5.
46. Cox DR. Regression models and life-tables. *J R Stat Soc Ser B* 1972; 34: 187–202.
47. Faraggi D and Simon R. A neural network model for survival data. *Stat Med* 1995; 14: 73–82.
48. Harrell FE. Evaluating the yield of medical tests. *JAMA J Am Med Assoc* 1982; 247: 2543.
49. Violan C, Foguet-Boreu Q, Flores-mateo G, et al. Prevalence, determinants and patterns of multimorbidity in primary care: a systematic review of observational studies. *PLoS One* 2014; 9: e102149.
50. Boyd CM and Fortin M. Future of multimorbidity research: how should understanding of multimorbidity inform health system design? *Public Health Rev* 2010; 32: 451–474.
51. Schiøtz ML, Stockmarr A, Høst D, et al. Social disparities in the prevalence of multimorbidity - A register-based population study. *BMC Public Health* 2017; 17: 422.
52. Dugravot A, Fayosse A, Dumurgier J, et al. Social inequalities in multimorbidity, frailty, disability, and transitions to mortality: a 24-year follow-up of the Whitehall II cohort study. *Lancet Public Heal* 2020; 5: e42–e50.
53. Schäfer I, Hansen H, Schon G, et al. The influence of age, gender and socio-economic status on multimorbidity patterns in primary care. First results from the multicare cohort study. *BMC Health Serv Res* 2012; 12: 89.
54. Chen Y and Xu R. Mining cancer-specific disease comorbidities from a large observational health database. *Cancer Inform* 2014; 13(s1): CIN.S13893.
55. Hernández B, Reilly RB and Kenny RA. Investigation of multimorbidity and prevalent disease combinations in older Irish adults using network analysis and association rules. *Sci Rep* 2019; 9: 14567.
56. Abdalla SM, Yu S and Galea S. Trends in cardiovascular disease prevalence by income level in the United States. *JAMA Netw Open* 2020; 3: e2018150.
57. Yoon SS, Dillon CF, Illoh K, et al. Trends in the prevalence of coronary heart disease in the U.S.: national health and nutrition examination survey, 2001–2012. *Am J Prev Med* 2016; 51: 437–445.
58. Benjamin EJ, Muntner P, Alonso A, et al. Heart disease and stroke statistics—2019 update: a report from the American heart association. *Circulation* 2019; 139: e56–e528.
59. Lampe FC, Morris RW, Whincup PH, et al. Is the prevalence of coronary heart disease falling in British men? *Heart* 2001; 86: 499–505.
60. Boyle JP, Honeycutt AA, Narayan KM, et al. Projection of diabetes burden through 2050: impact of changing demography and disease prevalence in the U.S. *Diabetes Care* 2001; 24: 1936–1940.
61. Boulet LP. Influence of comorbid conditions on asthma. *Eur Respir J* 2009; 33: 897–906.
62. Bardin PG, Rangaswamy J and Yo SW. Managing comorbid conditions in severe asthma. *Med J Aust* 2018; 209: S11.e3–S17.

63. Lago RM, Singh PP and Nesto RW. Diabetes and hypertension. *Nature clin pract endocrinol metab* 2007; 3: 667.
64. Long AN and Dagogo-Jack S. Comorbidities of diabetes and hypertension: mechanisms and approach to target organ protection. *J Clin Hyper* 2011; 13: 244–251.
65. de Boer IH, Bangalore S, Benetos A, et al. Diabetes and hypertension: a position statement by the American diabetes association. *Diabetes Care* 2017; 40: 1273–1284.
66. Centers for Disease Control and Prevention. *Coronary Artery Disease*, 2019, [https://www.cdc.gov/heartdisease/coronary\\_ad.htm](https://www.cdc.gov/heartdisease/coronary_ad.htm)
67. Rana JS, Nieuwdorp M, Jukema JW, et al. Cardiovascular metabolic syndrome - An interplay of, obesity, inflammation, diabetes and coronary heart disease. *Diabetes Obes Metab* 2007; 9: 218–232.
68. Rocca WA, Boyd CM, Grossardt BR, et al. Prevalence of multimorbidity in a geographically defined american population: patterns by age, sex, and race/ethnicity. *Mayo Clin Proc* 2014; 89: 1336–1349.
69. Maselli DJ and Hanania NA. Asthma COPD overlap: impact of associated comorbidities. *Pulm Pharmacol Ther* 2018; 52: 27–31.
70. Déruaz-Luyet A, N’Goran AA, Senn N, et al. Multimorbidity and patterns of chronic conditions in a primary care population in Switzerland: a cross-sectional study. *BMJ Open* 2017; 7: e013664.
71. Soley-Bori M, Bisquera A, Ashworth M, et al. Identifying multimorbidity clusters with the highest primary care use: 15 years of evidence from a multi-ethnic metropolitan population. *Br J Gen Pract* 2022; 72: e190–e198.
72. Chemey DZI, Repetto E, Wheeler DC, et al. Impact of cardio-renal-metabolic comorbidities on cardiovascular outcomes and mortality in Type 2 diabetes mellitus. *Am J Nephrol* 2020; 51: 74–82.
73. Arnold SV, Kosiborod M, Wang J, et al. Burden of cardio-renal-metabolic conditions in adults with type 2 diabetes within the diabetes collaborative registry. *Diabetes, Obes Metab* 2018; 20: 2000–2003.
74. Arnold SV, Hunt PR, Chen H, et al. Cardiovascular outcomes and mortality in type 2 diabetes with associated cardio-renal-metabolic comorbidities. *Diabetes* 2018; 67: 1582.
75. Lee SJ, Lindquist K, Segal MR, et al. Development and validation of a prognostic index for 4-year mortality in older adults. *JAMA* 2006; 295: 808.
76. Walter LC, Brand RJ, Counsell SR, et al. Development and validation of a prognostic index for 1-year mortality in older adults after hospitalization. *JAMA* 2001; 285: 2987.
77. Fortin M, Hudon C, Haggerty J, et al. Prevalence estimates of multimorbidity: A comparative study of two sources. *BMC Health Serv Res* 2010; 10: 111.
78. Fortin M, Stewart M, Poitras M-E, et al. A systematic review of prevalence studies on multimorbidity: toward a more uniform methodology. *Ann Fam Med* 2012; 10: 142–151.
79. Diederichs C, Berger K and Bartels DB. The measurement of multiple chronic diseases - a systematic review on existing multimorbidity indices. *Journals Gerontol - Ser A Biol Sci Med Sci* 2011; 66: 301–311.
80. The Diabetes Prevention Program (DPP) Research Group. The diabetes prevention program (DPP): description of lifestyle intervention. *Diabetes Care* 2002; 25: 2165–2171.