The Effect of Disease Co-occurrence Measurement on Multimorbidity Networks

by

Barret A. Monchka

A Thesis submitted to the Faculty of Graduate Studies of
The University of Manitoba
in partial fulfillment of the requirements of the degree of

MASTER OF SCIENCE

Department of Community Health Sciences
Max Rady College of Medicine
Rady Faculty of Health Sciences
University of Manitoba
Winnipeg

Abstract

Background: Network analysis, a technique for describing relationships, can provide insights into patterns of co-occurring chronic diseases. The effect that co-occurrence measurement has on disease network structure and resulting inferences has not been well studied.

Objectives: The research objectives were to (1) compare structural differences among chronic disease networks constructed from different co-occurrence measures, and (2) demonstrate how co-occurrences among three or more chronic diseases can be analyzed using network techniques.

Methods: A retrospective cohort study was conducted using four years of Manitoba administrative health data (2015/16 – 2018/19, 1.5 million individuals). Association rule mining was used to identify disease triads. Separate disease networks were constructed using seven co-occurrence measures: lift, relative risk, phi, Jaccard, cosine, Kulczynski, and joint prevalence. Influential diseases were identified using degree centrality and community detection was used to identify disease clusters. Community structure similarity was measured using the adjusted Rand index (ARI). Network edges were described using disease prevalence categorized as low (<1%), moderate (1% to <7%), and high (≥7%).

Results: Relative risk and lift highlighted co-occurrences between pairs of low prevalent diseases. Kulczynski emphasized relationships between conditions of high and low prevalence. Joint prevalence focused on highly prevalent conditions. Phi, Jaccard, and cosine emphasized associations with moderately prevalent conditions. Co-occurrence measurement differences significantly affected how disease clusters were defined, including the number of clusters identified. When limiting the number of edges to produce visually interpretable graphs, networks had significant dissimilarity in the percentage of co-occurrence relationships in common, and in their selection of the highest degree nodes.

Conclusion: Multimorbidity network analyses are sensitive to disease co-occurrence measurement. Co-occurrence measures should be selected considering research objectives and the prevalence relationships of greatest interest. Researchers should be cautious in their interpretation of findings from network analysis and should conduct sensitivity analyses using different co-occurrence measures. Many chronic diseases co-occur in groups of three or more and these higher-order associations can be visualized and analyzed using hypergraphs.

Acknowledgements

This contribution to scientific literature was made possible through the guidance and mentorship of Dr. Lisa Lix, who demonstrated her strong dedication to scholarship excellence as thesis advisor; and through the insightful and constructive feedback provided by advisory committee members Dr. Nathan Nickel and Dr. Carson Leung, who remarkably enriched this project with their respective expertise in epidemiology and computational methods.

This research was financially supported through a Frederick Banting and Charles Best Canada Graduate Scholarship from the Canadian Institutes of Health Research (CIHR), Research Manitoba Master's Studentship, University of Manitoba Graduate Fellowship, Research Manitoba and Rady Faculty of Health Sciences Graduate Entrance Scholarship, Visual and Automated Disease Analytics (VADA) Program Studentship, and a research fellowship provided by Dr. Lix.

The author acknowledges the Manitoba Centre for Health Policy for use of data contained in the Population Health Research Data Repository (Health Information Privacy Committee project # 2020/2021-01). The results and conclusions are those of the author and no official endorsement by the Manitoba Centre for Health Policy, Manitoba Health and Seniors Care, or other data providers is intended or should be inferred.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

Abbreviation	Definition
ACG	Adjusted Clinical Group
AIDS	Acquired immunodeficiency syndrome
AMI	Acute myocardial infarction
ARI	Adjusted Rand index
ARM	Association rule mining
COPD	Chronic obstructive pulmonary disease
EDC	Expanded Diagnostic Cluster
ENT	Ear, nose, and throat
ESRD	End-stage renal disease
FP-Growth	Frequent Pattern Growth algorithm
HIV	Human immunodeficiency virus
ICD	International Statistical Classification of Diseases and Related Health Problems
ICD-10-CA	International Statistical Classification of Diseases and Related Health Problems, 10th Revision, Canada
ICD-9-CM	International Statistical Classification of Diseases and Related Health Problems, 9th Revision, Clinical Modification
MEDC	Major Expanded Diagnostic Cluster
OR	Odds ratio
PAOH	Parallel Aggregated Ordered Hypergraph
RR	Relative risk
SARS	Severe acute respiratory syndrome
SCI	Salton Cosine Index
WHO	World Health Organization
ϕ	Pearson phi correlation coefficient
χ^2	Chi-square statistic

Chapter 1: Introduction

1.1 Background

Multimorbidity, the co-existence of two or more chronic health conditions within an individual, where none are considered more central than the others, is becoming increasingly common in Canada, as well as globally.[1,2] An aging population and increased life expectancy are two main drivers of the increasing prevalence of multiple chronic conditions in Canada.[1] Rising rates of behavioral risk factors, including physical inactivity, substance abuse, stress, and poor diet are also contributing to the rise in multimorbidity.[1,3] Those living with multiple chronic conditions tend to experience poorer quality of life, have increased disability and mortality, and face many challenges accessing healthcare services: conflicting medical advice, duplicative and unnecessary testing, drug interactions, and a heavy treatment burden.[1,4,5] Multimorbidity also places a strain on healthcare systems since individuals with multiple chronic conditions have higher healthcare utilization and costs.[6,7]

Network analysis, the study of relationships amongst connected entities, has been proposed as a method to shed new light on patterns of chronic disease in the population. Network analysis models disease co-occurrence using graph structures characterized by nodes (e.g., diseases) and connecting edges (i.e., relationships or interactions). Network edges may be directed, to include temporal disease progression information, or undirected; and weighted, to incorporate the strength of association, or unweighted. Several recent studies applied network analysis to electronic health data, to examine associations among co-occurring diseases at the population level.[8–20] Network analysis is appealing for chronic disease research, in part because of its reliance on graphical techniques to present disease associations, which can efficiently convey information in a non-technical manner to clinicians, patients, and decision makers. Network analysis also enables 1) the detection of important nodes or hubs, that is, diseases that are influential in a population or among a set of other diseases; 2) the identification of community structure, which represents clusters of highly-connected diseases; and 3) comparisons between population subgroups by contrasting subnetwork properties such as complexity measures.

Measuring disease association, or co-occurrence, is foundational for constructing the links that form the structure of disease networks. There are many co-occurrence measures available to choose from, and network analyses conducted to date have used a variety of different measures for constructing disease networks. The effect that the choice of co-occurrence measurement has on disease network structure and any resulting inferences has not been well studied. Although data mining techniques have been proposed for constructing disease networks based on associations of three or more diseases,[21] most network analyses construct disease networks using pairwise associations and few studies have incorporated knowledge from higher-order associations (i.e., ≥ 3 diseases). Network studies that extracted higher-order sets of co-occurring conditions did not incorporate all available information since only pairwise links were used to represent the higher-order associations.[20–30] Incorporating knowledge of higher-order disease combinations may provide additional insight useful for identifying clusters and central nodes.

Two recent systematic reviews found great variation in multimorbidity research methods, which could challenge the comparability of research findings[31,32] Research comparing different methodological approaches, for studying patterns of multimorbidity, has been recommended to improve study validity and generalizability.[32] Comparing techniques for constructing networks could aid in determining how different techniques affect our understanding of population-level chronic disease patterns. Since subgroup network comparisons and the identification of hubs and communities are three of the main components of network analysis, it is important to examine the effects that different disease co-occurrence methods have on network complexity, node centrality, and community structure. Comparing the effects of different disease co-occurrence methods could help develop guidelines for network analyses and direct future multimorbidity research.

1.2 Purpose and Objectives

The research purpose was to compare methods for measuring chronic disease co-occurrence in network analysis. The objectives were to (1) compare structural differences among chronic disease networks constructed from different co-occurrence measures, and (2) demonstrate how co-occurrences among three or more chronic diseases can be represented and analyzed using network techniques.

Chapter 2: Review of Literature

2.1 Disease Co-occurrence Measurement in Network Studies

Twenty-four studies were identified that used network techniques to analyze comorbidity and multimorbidity patterns in a variety of populations (Table 1, Table 2). These network analyses identified many known patterns of disease co-occurrence, as well as potentially novel disease associations for further investigation. Six different co-occurrence measures were used across the fourteen studies not employing association rule mining (Table 1), with the Pearson phi correlation coefficient (*n*=6, Equation 1), relative risk (*n*=3, Equation 2), and the odds ratio (*n*=3) being the most commonly used measures.

$$\phi = \frac{\mathrm{ad} - \mathrm{bc}}{\sqrt{(\mathrm{a} + \mathrm{b})(\mathrm{c} + \mathrm{d})(\mathrm{a} + \mathrm{c})(\mathrm{b} + \mathrm{d})}} \tag{1}$$

$$\mathrm{RR} = \frac{\mathrm{a}(\mathrm{c} + \mathrm{d})}{\mathrm{c}(\mathrm{a} + \mathrm{b})} \tag{2}$$

(relative risk of disease x if diagnosed with disease y)

given contingency table

		Disease x		
		Yes	No	Total
Disease y	Yes	a	b	(a + b)
	No	c	d	(c + d)
	Total	(a + c)	(b + d)	N

where N = total number of study participants

Hidalgo et al. stated the phi (ϕ) coefficient reliably measures associations between two diseases of similar prevalence (i.e. both highly prevalent or both rare), but is likely to underestimate associations between rare and prevalent diseases; whereas relative risk is stated as underestimating associations between two highly prevalent diseases, and overestimating

Table 1. Characteristics of studies using network analysis to analyze patterns of co-occurring disease.

Study (year)	Study type	Index condition(s)	Data Source		Disease co-occurrence measure(s) (effect size threshold)	Association set size
			Type	Country		
Chmiel et al. (2014)[8]	Multimorbidity		Administrative health data	Austria	Adjusted phi (not stated)	2
Divo et al. (2015)[33]	Comorbidity	COPD	Study	Spain, United States	Phi (not stated)	2
Davis & Chawla (2011)[12]	Multimorbidity		Electronic medical/health records	United States	Mutual information weighting (not stated)	2
Duarte et al. (2017)[9]	Comorbidity	Cancer, cardiovascular disease	Administrative health data	United Kingdom	Odds ratio (> 1)	2
Hanauer & Ramakrishnan (2013)[10]	Multimorbidity		Electronic medical/health records	United States	Odds ratio (≥300, ≥800)	2
Hidalgo et al. (2009)[11]	Multimorbidity		Administrative health data	United States	Phi (> 0.06), relative risk (> 20)	2
Jeong et al. (2017)[13]	Multimorbidity		Administrative health data	South Korea	Relative risk (> 4)	2
Jiang et al. (2018)[14]	Multimorbidity		Administrative health data	Taiwan	Phi (not stated)	2
Kalgotra et al. (2017)[15]	Multimorbidity		Electronic medical/health records	United States	Salton Cosine Index (≥ 0.04)	2
Khan et al. (2018)[16]	Comorbidity	Type 2 diabetes	Administrative health data	Australia	Frequency (≥ 1)	2
Kim et al. (2016)[17]	Multimorbidity		Administrative health data	South Korea	Odds ratio (> 5)	2
Lai (2016)[18]	Comorbidity	HIV/AIDS	Administrative health data	Taiwan	Phi (> 0.06)	2
Moni & Liò (2014)[19]	Comorbidity	HIV-1, SARS	Administrative health data	United States	Phi (≥ 0.06), relative risk (≥ 10, ≥ 20, ≥ 100)	2
Schäfer et al. (2014)[20]	Multimorbidity		Administrative health data	Germany	Observed-to-expected ratio (≥ 2)	3

Note: Studies using association rule mining were excluded. AIDS = acquired immunodeficiency syndrome, COPD = chronic obstructive pulmonary disease, HIV = human immunodeficiency virus, SARS = severe acute respiratory syndrome.

associations between two rare diseases.[11] Considering this, small estimates of correlation may indicate truly weak associations between two diseases, or instead be the result of large differences in prevalence estimates.[8] Hidalgo et al. compared disease networks constructed using RR and ϕ and found the network constructed with RR contained a higher number of low prevalence conditions, while the network constructed using ϕ contained a greater number of highly prevalent conditions.[11] Links between disease nodes and the resulting network modules (i.e., community structure) differed between the two networks: the ϕ-based network had more connections between different disease categories, while the RR network contained more connections within disease categories.[11] This suggests the choice of association measure can impact inferences made using network analysis. Hidalgo et al. indicated that each association measure provided a different representation of a disease network, and did not recommend one over the other.[11] Chmiel et al. applied an adjustment for the bias inherent in the ϕ coefficient by dividing the estimate, between a rare and prevalent disease, by the typical correlation strength for the rare disease.[8]

Other alternatives to ϕ and RR have also been used in network analyses of co-occurring disease. Kalgotra et al. used the Salton Cosine Index (SCI) because they suggested it is not influenced by the number of observations, unlike the chi-square statistic (χ^2) which is affected by sample size.[15] Davis and Chawla used mutual information weighting, which compares the joint probability of two diseases with the product of their marginal probabilities, to minimize bias based on disease prevalence when constructing their disease network.[12] Schafer et al. measured association using observed-expected ratios, and extended their analysis beyond co-occurring disease pairs to disease triads.[20] None of the reviewed network analyses estimated disease co-occurrence using a null-invariant measure. Unlike ϕ and RR, associations between diseases measured using null-invariant measures are not affected by increasing the number of individuals containing none of the diseases under inspection. It has been suggested that null-invariant measures may be more appropriate for association analysis performed in large databases that contain a large proportion of null transactions (observations that do not contain any of the events of interest).[34,35] This suggests null-invariant measures of association may be applicable for disease co-occurrence studies since disease status matrices contain mainly null values. Jaccard (Equation 3), cosine (Equation 4), and Kulczynski (Equation 5) are three null-invariant measures that differ in the types of relationships they assign higher weights towards.[34] Jaccard tends to prefer relationships between two events of similar frequency, Kulczynski assigns higher weights towards skewed relationships (i.e., between frequent and rare events), and cosine tends to compromise between these two approaches.[36]

$$\mathrm{Jaccard}(\mathrm{X},\mathrm{Y}) = \frac{\mathrm{P}(\mathrm{X} \cap \mathrm{Y})}{\mathrm{P}(\mathrm{X}) + \mathrm{P}(\mathrm{Y}) - P(\mathrm{X} \cap \mathrm{Y})} \tag{3}$$

where X and Y are itemsets (i.e., sets of disease categories)

$$\mathrm{cosine}(\mathrm{X},\mathrm{Y}) = \sqrt{\mathrm{P}(\mathrm{X}|\mathrm{Y})\mathrm{P}(\mathrm{Y}|\mathrm{X})} \tag{4}$$

$$\mathrm{Kulczynski}(\mathrm{X},\mathrm{Y}) = \frac{1}{2}(\mathrm{P}(\mathrm{X}|\mathrm{Y}) + \mathrm{P}(\mathrm{Y}|\mathrm{X})) \tag{5}$$

Since multimorbidity is modified by sociodemographic variables such as age and sex, disease associations that are adjusted for these factors may be beneficial. Duarte et al. adjusted for demographic and lifestyle factors in a logistic regression model to produce adjusted odds ratios (ORs) of disease pairings.[9] Other studies used stratification to create separate networks for demographic factors, such as sex.[15,20] Divo et al. used a case-control study design and stratified results by creating separate disease networks based on the presence of an index condition.[33] However, most network analyses used crude measures of association with no adjustment for confounders or other covariates.

Due to the large number of statistical tests of association that are typically performed in a network analysis, there is an increased likelihood of obtaining statistically significant association estimates for disease patterns with little clinical or practical significance. Researchers may wish to reduce the number of associations by using effect size cut-offs, adjusting the nominal level of statistical significance (i.e., α), applying family-wise error adjustment (e.g., Bonferroni correction), or by decreasing the false discovery rate (e.g., Benjamini-Hochberg Procedure[37]). Celli et al. used a strict p-value cut-off of 0.01 to account for the increased Type I error rate,[38] while Kim et al. used a Bonferroni correction ($p < 1.38 \times 10^{-7}$) when performing χ^2 tests of odds ratios.[17] The downside to using conservative p-value cut-offs, family-wise error adjustments, or false discovery rates is the increased possibility of discarding interesting associations;[17] and even with a multiple comparisons adjustment, statistically significant disease co-occurrences are not necessarily clinically or practically significant.[39]

2.2 Association Rule Mining for Extracting Disease Co-occurrence Patterns

Association rule mining is a data mining technique for extracting interesting patterns among dataset variables. In comparison to pairwise statistical association analysis, ARM offers the potential to discover associations among higher-order sets (i.e., ≥ 3 diseases). Chen et al. suggested ARM is less susceptible to the biases inherent in RR and ϕ when there are large differences in prevalence for the disease pairs being considered.[24]

ARM consists of two main steps: (1) find all frequent itemsets of interest (e.g., the most frequently observed co-occurrence relationships) using a frequent pattern mining algorithm (e.g., Apriori[40]), and (2) generate association rules from the mined frequent itemsets. Association rules are directional, consisting of an antecedent and a consequent, and represent relationships between two sets of variables. In the case of analyzing disease co-occurrence, the antecedent and consequent are sets of diagnosis codes, or disease categories. For example, the association rule {x, y} → {z} represents the tendency of individuals diagnosed with disease x and disease y (antecedent) to also be diagnosed with disease z (consequent). Although association rules are directional, they do not imply causality but instead represent co-occurrence relationships between the antecedent and the consequent.[41]

The strength of an association rule has traditionally been determined by its support and confidence measurements. Support is defined as the proportion of observations (i.e., individuals) that contain all items (i.e., diagnosis codes) appearing in both the antecedent (X) and the consequent (Y) itemsets. This is equivalent to measuring the joint probability of certain diagnosis codes occurring within an individual's health record (Equation 6).[41]

$$\mathrm{support}(\mathrm{X} \Longrightarrow \mathrm{Y}) \;=\; \mathrm{P}(\mathrm{X} \cup \mathrm{Y}) \tag{6}$$

In epidemiological terminology, support is synonymous with the joint prevalence of all diseases listed in an association rule. The confidence measure represents the proportion of observations containing the antecedent, which also contain the consequent. Confidence is defined as the conditional probability of the consequent, given the antecedent (Equation 7);[41] which is synonymous to the prevalence of a set of comorbidities (X) among individuals in a population with a specific set of index conditions (Y).

$$\mathrm{confidence}(\mathrm{X} \Longrightarrow \mathrm{Y}) \;=\; \mathrm{P}(\mathrm{X}|\mathrm{Y}) \tag{7}$$

Higher confidence values indicate a higher likelihood for observations to include all the items in an association rule.

Frequent pattern mining algorithms, such as the Apriori algorithm, require a user-defined support threshold to be set. Only itemsets with a frequency greater than the minimum support threshold are included in the extracted results. If the minimum support threshold is set too high, strongly associated items that occur infrequently may be excluded.[42] This situation is known as the rare item problem and could be reduced if a low support threshold is chosen.[42] However, lower support thresholds may generate too many uninteresting associations.[41] Association rule mining also requires a minimum confidence threshold to be supplied, and generated association rules are only considered interesting if their confidence value is greater than this minimum. Minimum support and confidence thresholds are not trivial to define; and their choice should be based on the length of the dataset, sparseness of the data, and domain knowledge.[43] Support and confidence measurements alone are unable to adequately distinguish between interesting and non-interesting associations.[44] Using only support and confidence measures to discard uninteresting patterns can lead to the inclusion of uninteresting results and the rejection of practically significant patterns.[41]

Correlation measures, such as the lift measure, can be used to improve upon the classical support-confidence framework and filter out misleading strong associations using the concept of probabilistic independence.[41,44] Lift is defined as the ratio of the support of an association rule to what would be expected under statistical independence (Equation 8).

$$\mathrm{lift}(\mathrm{X} \Longrightarrow \mathrm{Y}) \;= \frac{\mathrm{P}(\mathrm{X} \cup \mathrm{Y})}{\mathrm{P}(\mathrm{X}) * \mathrm{P}(\mathrm{Y})} \tag{8}$$

The range of possible lift values differs between association rules and depends upon the support of the antecedent and consequent.[45]

Traditionally in data mining, tests of statistical significance are not used when determining the “interestingness” of an association rule. As a result, there is no assumed underlying probability distribution in the classical use of the support, confidence, or lift interestingness measures. This is in contrast to the use of statistical significance testing seen with other association measures such as RR and ϕ, which assume an underlying probability distribution for the test statistic. However, some health-related studies have used the chi-square statistic (Equation 9) to assess the statistical significance of association rules.[46,47]

$$\chi^2(\mathrm{X} \Rightarrow \mathrm{Y}) = \mathrm{n}(\mathrm{lift}-1)^2 \frac{\mathrm{supp} * \mathrm{conf}}{(\mathrm{conf}-\mathrm{supp})(\mathrm{lift}-\mathrm{conf})} \tag{9}$$

$$\begin{aligned} \text{where } \mathrm{sup} &= \mathrm{support}(\mathrm{X} \Rightarrow \mathrm{Y}), \\ \mathrm{conf} &= \mathrm{confidence}(\mathrm{X} \Rightarrow \mathrm{Y}), \\ \mathrm{lift} &= \mathrm{lift}(\mathrm{X} \Rightarrow \mathrm{Y}) \end{aligned}$$

Several studies have used association rule mining to analyze patterns of disease co-occurrence (Table 2). Of the twenty studies identified, the majority (*n*=13, 65%) analyzed comorbidities in relation to an index condition; while multimorbidity was investigated in seven (35%) of the studies. The majority of the studies used U.S.-based data (*n*=8, 40%), while only one study used Canadian data.

Most of the ARM-based studies defined support and confidence thresholds to limit the number of association rules. All of the studies that defined support thresholds did so at a low level (≤10%), with a range of 0.1% to 10%; while confidence thresholds varied greatly among the studies, ranging from 0.5% to 90%. Held et al. left support unbounded;[25] while Hernandez et al. and Shen et al. left support and confidence unbounded and relied on lift cut-offs to filter potentially uninteresting association rules.[26,48] 35% (*n*=7) of the studies used lift to either rank the mined associations or to exclude association rules. Three studies required association rules to have lift > 1, while one study used lift ≥ 2. Hernandez et al. excluded association rules having standardized lift values ≤ 0.2.[26] Apriori was the most commonly used frequent pattern mining algorithm (*n*=10), while four studies used the Frequent Pattern Growth algorithm (FP-Growth)[49] and one was based on the Eclat algorithm.[50]

Ten (50%) of the identified studies used network techniques to visualize or analyze the disease co-occurrence relationships obtained using ARM. Four of these network analyses studied multimorbidity, but none were conducted using population-based diagnostic health records. Seven of the network analyses extracted higher-order associations (i.e., ≥ 3 diseases); however,

Table 2. Characteristics of studies using association rule mining to analyze disease co-occurrence patterns.

Study (year)	Study type	Index condition(s)	Network analysis	Data Source		Frequent pattern mining algorithm[1]	Interestingness measure(s) (thresholds)[2]	Maximum itemset size[3]
				Type	Country			
Chen & Xu (2014)[23]	Comorbidity	Cancer	✓	Adverse event reports	United States	FP-growth	Support (N≥5), confidence (>10%)	3
Chen et al. (2015)[24]	Comorbidity	Colorectal cancer, obesity	✓	Adverse event reports	United States	Not specified	Confidence (>50%)	Not specified
Held et al. (2015)[25]	Comorbidity, multimorbidity	Frailty, falls	✓	Study	Australia	Eclat	Support (unbounded), confidence (>10%), lift (≥2)	Unbounded
Hernandez et al. (2019)[26]	Multimorbidity		✓	Study	Ireland	Not specified	Support (unbounded), confidence (unbounded), standardized lift (>0.2)	3
Ho et al. (2019)[51]	Multimorbidity			Electronic medical/health records	United States	Apriori	Support (>0.1%), confidence (>5%)	3
Kang'ethe & Wagacha (2014)[52]	Multimorbidity			Electronic medical/health records	United States	Apriori	Support (varied), confidence (varied)	Not specified
Kim et al. (2012)[53]	Comorbidity	Type 2 Diabetes Mellitus		Electronic medical/health records	South Korea	Apriori	Support (>3%), confidence (>5%)	3
Kim & Myoung (2018)[27]	Comorbidity	Attention-deficit Hyperactivity Disorder	✓	Administrative health data	South Korea	Apriori	Support (≥1%), confidence (≥50%)	3
Madlock-Brown & Reynolds (2019)[54]	Comorbidity	Obesity		Electronic medical/health records	United States	FP-growth	Support (>10%), confidence (>60%)	3
Nassar & Richter (2018)[55]	Comorbidity	Gastroparesis		Electronic medical/health records	United States	Apriori	Not specified	2
Peng et al. (2018)[56]	Data quality			Administrative health data	Canada	Apriori	Support (≥0.19%), confidence (≥50%)	5
Shen et al. (2017)[48]	Comorbidity	Borderline personality disorder		Administrative health data	Taiwan	Apriori	Support (0%), confidence (0%), lift (>1)	4
Shin et al. (2010)[28]	Comorbidity	Essential hypertension	✓	Electronic medical/health records	South Korea	Apriori	Support (≥5%), confidence (≥15%), lift(unbounded)	3
Tai & Chiu (2009)[21]	Comorbidity	Attention-deficit Hyperactivity Disorder	✓	Administrative health data	Taiwan	Apriori	Support (>4%), confidence (>90%)	3
Valent et al. (2013)[57]	Comorbidity	Diabetes Mellitus		Administrative health data	Italy	Not specified	Support (>0.5%), confidence (>5%)	3
Wang et al. (2019)[58]	Comorbidity	Mental disorders		Administrative health data	Taiwan	Apriori	Support (>2%)	3
Yao et al. (2019)[59]	Multimorbidity			Study	China	Not specified	Support (>2%), confidence (>10%), lift (>1)	2
Zemedikun et al. (2018)[29]	Multimorbidity		✓	Study	United Kingdom	Not specified	Support (not specified), confidence (not specified), lift (not specified)	3
Zheng & Xu (2018)[30]	Multimorbidity		✓	Adverse event reports	United States	FP-growth	Support (>12), confidence (>0.5)	Not specified
Zheng & Xu (2019)[22]	Comorbidity	Alzheimer's disease	✓	Adverse event reports	United States	FP-growth	Support (>12), lift (>1)	Not specified

1. Computational algorithm for extracting frequently co-occurring disease sets; 2. Measure of association rule importance (minimum value cut-off); 3. Maximum number of frequently co-occurring diseases extracted

all of these studies used pairwise edges to represent these relationships. None of the ARM-based network studies used hypergraph structures, generalizations of graphs where edges can connect any number of nodes, to represent associations amongst higher-order disease sets.

2.3 Higher-order Disease Associations

Network data is commonly modeled with pairwise links to indicate relationships between pairs of entities.[60] These relationships are visually expressed using binary edges, which connect pair of nodes within graph structures. However, many real-world phenomena contain relationships between three or more entities and traditional binary networks are unable to fully model the complexity of these real-world systems.[60,61] Network analysis limited to pairwise associations may not identify the desired community structure and nodes of importance in complex systems that feature many higher-order co-occurrence relationships (i.e., associations amongst three or more entities). Fotouhi et al. suggest analyzing associations among higher-order sets, in comparison to pairwise associations, in order to more accurately capture disease progression in network analyses.[62] Doulis suggested higher-order disease associations have the potential to provide additional insight into disease association and progression, and proposed studying the effects of higher-order disease groups in future work.[63]

Hypergraphs are generalizations of graphs that are not restricted to pairwise links, and support the modeling of higher-order co-occurrence relationships. Edges in hypergraphs, known as hyperedges, are able to link any number of network nodes; and are commonly visualized using coloured bounding containers, containing the nodes they link together. A hypergraph (H) is formally defined as a pair $H = (V, E)$ containing a set of vertices (V) and a set of hyperedges (E); while a hyperedge is defined by the set of vertices that it links (i.e., $E_1 = \{v_1, v_2, v_3\}$). Unlike edges in traditional graphs, hyperedges are not restricted to a set of only two nodes. Alternative visual representations include the use of non-binary edges, capable of connecting any number of nodes (i.e., one-to-many network edges); and the Parallel Aggregated Ordered Hypergraph (PAOH) visualization, a figure that visually represents hyperedges using vertical lines.[64] Hypergraphs can be analyzed using standard pairwise graphs if converted to their bipartite representations, where a hyperedge is represented by an additional node that links all of its respective vertices.[65]

Although hypergraphs are able to represent complex systems, most research using network techniques have continued to use pairwise networks. Few health studies have modeled higher-order interactions among network entities using hypergraphs.[60] A select number of studies employed hypergraph structures to analyze human disease;[66–68] however, no known studies used hypergraphs to model disease co-occurrence and instead modeled these multi-way relationships using pairwise graphs. Belyi et al. used the bipartite representation of a hypergraph to model higher-order combinations of prescription drugs frequently taken together.[69] However, using a bipartite graph artificially increases node and edge counts, alters network structure, and hampers visual interpretations of networks. The addition of nodes to represent hyperedges may also adversely affect community detection and the identification of central nodes.

A substantial percentage of Canadians are living with three or more chronic conditions,[70] and several multimorbidity studies identified frequent patterns of three or more co-occurring diseases in their study populations.[31] For example, a U.S.-based study by Majumdar et al. found the disease triad of diabetes, hypertension and hyperlipidaemia to commonly occur in their study population with a prevalence of 10%.[71] Network analyses using hypergraphs are able to model disease triads and larger combinations of co-occurring conditions, and incorporate that additional knowledge into the analysis.

Chapter 3: Methods

3.1 Study Design and Data Source

This retrospective cohort study was conducted using four fiscal years (April 1, 2015 – March 31, 2019) of administrative health data from the Manitoba Population Health Research Data Repository at the Manitoba Centre for Health Policy. Data sources were linked using a unique personal health identification number. The Health Research Ethics Board for the University of Manitoba approved this study and approval for data access was provided by the Health Information Privacy Committee for Manitoba Health and Seniors Care.

Study data sources included the Manitoba Health Services Insurance Plan Registry (Population Registry), the Hospital Abstracts Database, and the Medical Services Database. The Population Registry stores data on health care coverage for all insured Manitobans, and was used to determine eligibility for inclusion in this study. The Registry also includes demographic information (e.g., age and sex), which was used to characterize the study cohort and stratify the analyses. Chronic disease information was obtained from inpatient hospital discharge abstracts and billing claims from ambulatory encounters.

The Hospital Abstracts Database contains information on discharges from hospitals in Manitoba. Diagnoses within hospital discharge abstracts are coded using the International Statistical Classification of Diseases and Related Health Problems, 10th Revision with Canadian Enhancements (ICD-10-CA), since April 1, 2004. The Medical Services Database contains information on services provided in physician offices, and diagnoses are recorded using 5-digit ICD-9-CM codes since April 1, 2015. The 4-year study period (April 1, 2015 – March 31, 2019) was chosen to maximize diagnostic precision, since Medical Services diagnoses were recorded using only 3-digit ICD-9-CM codes prior to April 1, 2015.

3.2 Cohort

The study cohort included all Manitoba residents with complete or partial Manitoba Health insurance coverage during the 4-year study period (April 1, 2015 – March 31, 2019). Individuals entered the study on April 1, 2015 or the date that coverage started, and were followed until the end of the study period or until their insurance coverage ceased due to death, moving away from Manitoba, or other reasons. Chronic disease data obtained in subsequent coverage periods, for individuals that lost and later re-gained Manitoba Health insurance coverage, were included in the analysis.

Males with female-specific conditions and females with male-specific conditions were excluded since the presence of this inconsistency suggests either errors in diagnosis or demographic coding. Specifically, males were excluded if they were assigned a diagnosis of endometriosis; malignant neoplasms of the cervix, uterus, or ovary; or other female gynecologic conditions. Females were excluded if they recorded diagnoses of prostatitis, prostatic hypertrophy, malignant neoplasms of the prostate, or other male genital disease.

Since disease networks were formed from disease co-occurrence relationships, the network analysis was limited to individuals with diagnoses for at least two chronic conditions in the study observation period.

3.3 Disease Ascertainment

Chronic diseases were ascertained using diagnoses identified from inpatient discharge records in the Hospital Abstracts Database, and from physician visit records in the Medical Services Database. Surgeries recorded in both data sources were also included. Prenatal and pregnancy-related records were excluded to minimize overstating disease co-occurrence among females. A single diagnosis code was used to ascertain whether an individual was considered as having a specified condition in the study observation period. Individual diagnosis codes were grouped using two different methods: 1) into 31 categories based on the Elixhauser[72] comorbidity index (Appendix A, Appendix B), and 2) grouped into 201 Expanded Diagnostic Clusters (EDC) and 27 higher-level Major Expanded Diagnostic Clusters (MEDC) of the Johns Hopkins Adjusted Clinical Group (ACG) System (Appendix C).[73] Diagnoses were loaded into the Johns Hopkins System as World Health Organization (WHO) ICD-9 or ICD-10 codes. 5-digit ICD-10-CA codes from the Hospital Abstracts Database were truncated to the first four digits to improve compatibility with the Johns Hopkins System, which supports the WHO ICD system but not the Canadian revision. There were a total of 49 unique Canadian-specific ICD-10-CA codes relevant to chronic disease status that were not captured by the Johns Hopkins System. These 49 Canadian-specific diagnosis codes were first translated to WHO ICD-10 codes for inclusion. 17 additional Canadian-specific ICD-10-CA codes were not captured; however they were irrelevant to disease status since they indicated location of occurrence or activity engaged in during occurrence. Chronic conditions classified as separate EDC categories based on severity or presence of complications were combined into single disease categories including asthma with or without asthmaticus, hypertension with or without complications, type 1 diabetes with or without complications, and type 2 diabetes with or without complications. As well, 25 EDC categories that were non-descriptive, or referred to non-chronic medical conditions or to the neonatal period were removed from the analysis (Appendix C). Two categories indicating severity of malignant neoplasms, already classified elsewhere, were also excluded. Since co-occurrences with frequencies less than 15 were excluded from the association analysis to minimize statistical errors, seven EDC categories with low frequencies were removed: heart murmur, lymphadenopathy, thrombophlebitis, tuberculosis infection, sinusitis, other inflammatory conditions of skin, and other female gynecologic conditions. After a total of 34 EDC categories were excluded, 167 EDC categories remained for the network analysis.

3.4 Disease Co-occurrence Measurement

Disease co-occurrence was defined as two or more conditions recorded at any time during the 4-year study observation period, for the same individual. Disease association was measured using seven different co-occurrence measures: joint prevalence, relative risk (RR), phi (ϕ), lift, cosine, Jaccard, and Kulczynski.[36,74–76] Phi and relative risk are two of the most commonly used measures in disease network analysis, while lift is commonly used in conjunction with association rule mining. Cosine, Jaccard, and Kulczynski are null-invariant measures commonly recommended for sparse data such as disease status datasets. Joint prevalence was included due to its ease of interpretation. Disease co-occurrence was measured for the entire multimorbidity cohort, as well as for males and females separately. Statistical significance was assessed using the chi-square test when expected frequencies were greater than five, while Fisher's exact test was used when the chi-square assumption did not hold. Associations that were not statistically significant using α=0.01 were excluded. Since the focus of our study was on co-occurring disease, the analysis was limited to positive associations, and negative correlations and protective associations were excluded. Since RR is an asymmetric measure of association, the maximum of the two RR measures was used.

The association analysis was limited to disease dyads and triads, while associations among four or more diseases were excluded. The Apriori[40] algorithm was used to extract associations amongst sets of three co-occurring conditions. Minimum joint frequency (called support in association rule mining) was limited to 15 to minimize statistical errors, and the minimum confidence parameter of association rule mining was left unbounded. Data preprocessing and disease ascertainment was conducted using SAS, while R and the arules[77] package (v1.6-7) was used to perform the association analysis.

3.5 Covariates

The study cohort was characterized by age, sex, number of chronic conditions (based on the Johns Hopkins ACG System), residence location (urban or rural), socioeconomic status, and healthcare utilization. Since patterns of chronic disease differ by sex, separate disease networks were constructed for males and females.

The most recent demographic information submitted to Manitoba Health was assumed correct: birthdate and sex were extracted from the most recent insurance coverage period, while socioeconomic and urban/rural status were based on the latest residence recorded during the study period. Age was calculated at exit date (i.e., the study index date) and categorized as <20, 20-39, 40-59, 60+. Income quintile was calculated using the most recent available Canadian Census data (2016) and was based on residence location at the study index date. Hospital utilization was measured in binary format indicating whether an individual had at least one inpatient hospitalization during the 12 months prior to the study index date. Physician utilization was defined as the number of ambulatory visits recorded during the 12 months prior to the study index date. Prenatal and pregnancy diagnosis codes were excluded from hospital and physician utilization measures.

3.6 Network Analysis

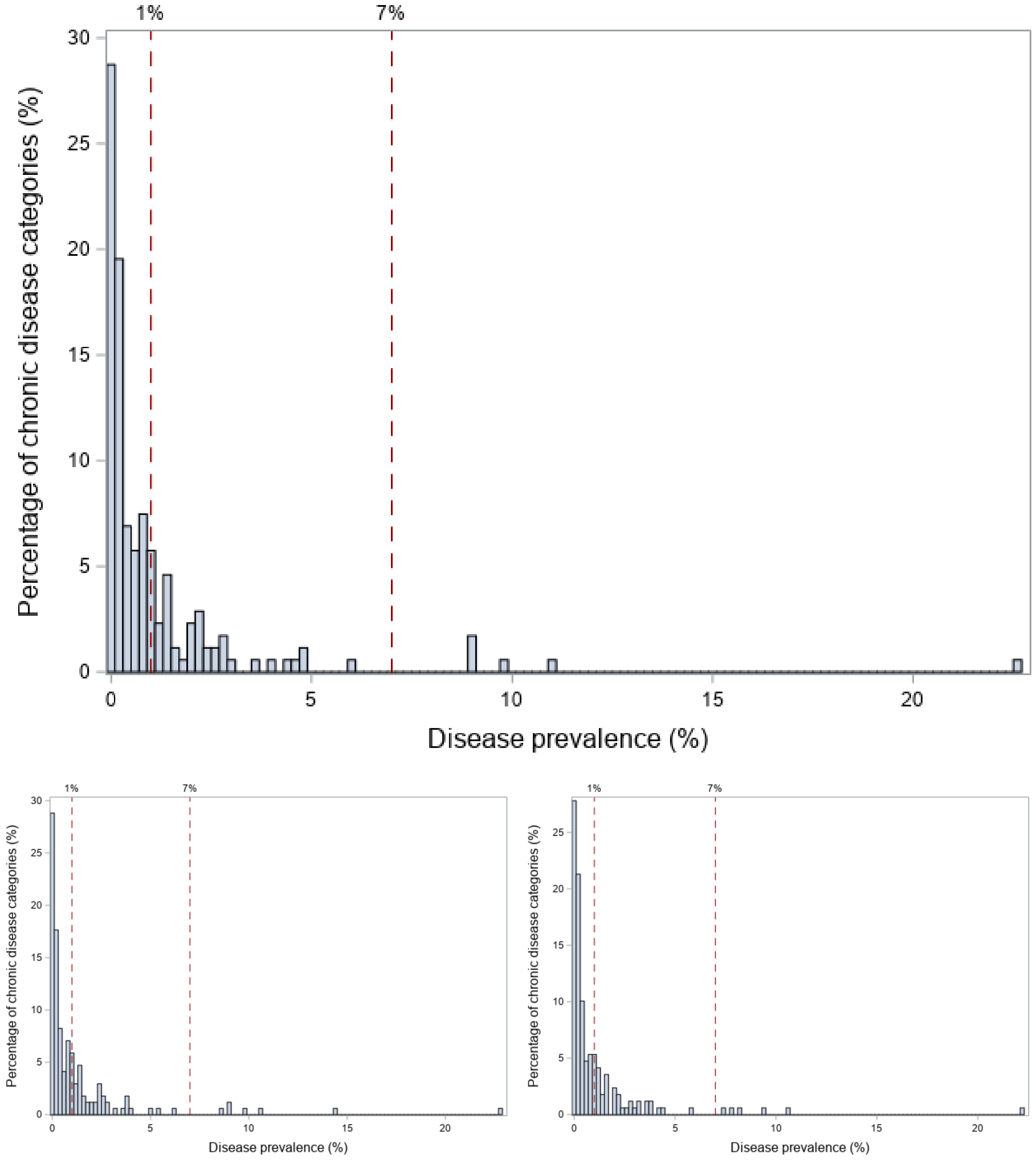
Weighted, undirected pairwise disease networks were separately constructed using the seven disease co-occurrence measures, and separately for Johns Hopkins EDC and Elixhauser[72] disease categories. Hypergraph structures were constructed using both pairwise and triad associations, and separately for each disease co-occurrence measure. Pairwise networks and hypergraphs based on EDC categories were further stratified by the number of associations (i.e., edges) included: all associations, strongest 50 percent of associations (i.e., highest effect size), and the strongest 200 associations. Networks were limited to the strongest 200 associations to examine differences in less complex networks that have higher visual interpretability, while the strongest 50 percent cut-off was chosen to examine how network similarity changes as a larger number of associations are included. Effect size estimates were used as edge weights and were bounded between 0 and 1 for networks measured using phi, Jaccard, cosine, and Kulczynski association measures; and unbounded for lift, relative risk, and joint prevalence.

Community structure in traditional pairwise networks was identified using a weighted and non-overlapping community detection algorithm developed by Blondel et al.[78] Hypergraph communities were identified using a community detection algorithm developed by Kamiński et al., using the SimpleHypergraphs.jl library in Julia.[79] Central nodes in pairwise and hypergraph network structures were identified using degree centrality (the number of co-occurrence relationships). Pairwise disease network visualizations were constrained to the strongest 200 associations, in order to produce visually interpretable network diagrams, and visualized using the Fruchterman-Reingold[80] force-directed network layout algorithm. Node size and node label text are proportional to disease prevalence, while edge thickness is proportional to effect size. Node and edge colours were assigned to indicate community structure. Pairwise network analysis was performed in Java using Gephi Toolkit (v0.9.2), and pairwise networks were visualized using Gephi (v0.9.2). The hypergraph constructed using phi was visualized using Python and HyperNetX[81] v1.0.2 (limited to the strongest 30 hyperedges), and as a Parallel Aggregated Ordered Hypergraph diagram using PAOHVis[64] v1.0.0 (limited to the strongest 100 hyperedges).

3.7 Evaluating and Comparing Disease Networks

Pairwise disease networks, constructed using different co-occurrence measures, were compared using network complexity measures, proportion of associations in common, and in terms of the joint prevalence and prevalence difference distributions of the network edges. Network edges were also compared by comparing categorized prevalence of co-occurring disease pairs. Based on the distribution across all 167 Johns Hopkins disease categories and sex-specific differences, disease prevalence was categorized as low (<1%), moderate (1 to <7%), and high (≥7%) (Figure 1). A sensitivity analysis was also performed by categorizing prevalence as low (<0.5%), moderate (0.5 to <5%), and high (≥5%). Global network properties used for characterizing and comparing networks included network density (the ratio of the number of edges present in a network to the number of possible edges between all node pairs), modularity (a measure of how well network nodes divide into communities), degree distribution, and node and edge counts. Important nodes, identified using degree centrality, were compared across networks by calculating the agreement percentage among the top 20 most central nodes. Community structure similarity was calculated using the adjusted Rand index (ARI) with the R package aricode (v1.0.0).[82] ARI measures the similarity between two data clusterings based on the number of pairs assigned to the same or different clusters, and adjusted for chance. ARI ranges from -1 to +1, with +1 indicating perfect similarity, 0 indicating cluster agreement is no better than random, and negative values indicating cluster similarity is worse than what would be expected for two random partitions. Pairwise networks and hypergraphs, constructed using the same co-occurrence measure, were contrasted with each other by comparing community structure and degree centrality distributions. The two network structures were also compared by extracting the binary relationships from the higher-order hyperedges and calculating the percentage of pairwise network associations that are represented in the respective hypergraph.

Figure 1. Prevalence distribution of chronic disease categories for all study participants (top), and females (bottom left) and males (bottom right) separately.



Note: Chronic diseases were ascertained using the Johns Hopkins ACG System. 1% and 7% vertical lines indicate cut-off points used for categorizing disease prevalence as low (<1%), moderate (1 to <7%), and high (≥7%).

Chapter 4: Results

4.1 Cohort

4.1.1 Demographic and Healthcare Utilization Characteristics

Out of 1,510,678 Manitoba residents with Manitoba Health insurance coverage between fiscal years 2015/16 and 2018/19, 610,427 (40.4%) had no chronic disease diagnosis recorded, 282,340 (18.7%) recorded a single chronic condition diagnosis, and 617,911 (40.9%) had two or more chronic condition diagnoses and were included in the network analysis (Table 3, Appendix E). Fifteen individuals recorded sex-specific diagnoses that were inconsistent with their Manitoba Health Insurance Registry record and were excluded from the study (Figure 2). The median age of individuals with multimorbidity was considerably higher (57 years, Q1-Q3: 41-70) than individuals with one chronic condition (33 years, Q1-Q3: 18-49) or without any chronic disease (24, Q1-Q3: 11-37). There were a higher percentage of females (54.1%) and urban residents (64.1%) with multimorbidity than without (47.1% female, 61.3% urban). There were only minor differences in the distribution of socioeconomic status (income quintile) between those with and without multimorbidity. Individuals with a diagnosed chronic disease were higher users of physician services: 86.8% (*n*=245,091) of individuals living with one chronic condition and 97.4% (*n*=601,899) of those with multimorbidity recorded an ambulatory visit during the last year of follow-up; while 59.2% (*n*=361,628) of individuals without a diagnosed chronic disease had at least one ambulatory encounter. The percentage of individuals with a recorded inpatient hospitalization during the last 12 months of follow-up was significantly higher for those with multimorbidity (13.6%, *n*=83,934) compared with individuals without multimorbidity (4.1%, *n*=36,619).

Table 3. Demographic and chronic disease characteristics of Manitoba residents with multimorbidity (*n*=617,911), 2015/16-2018/19.

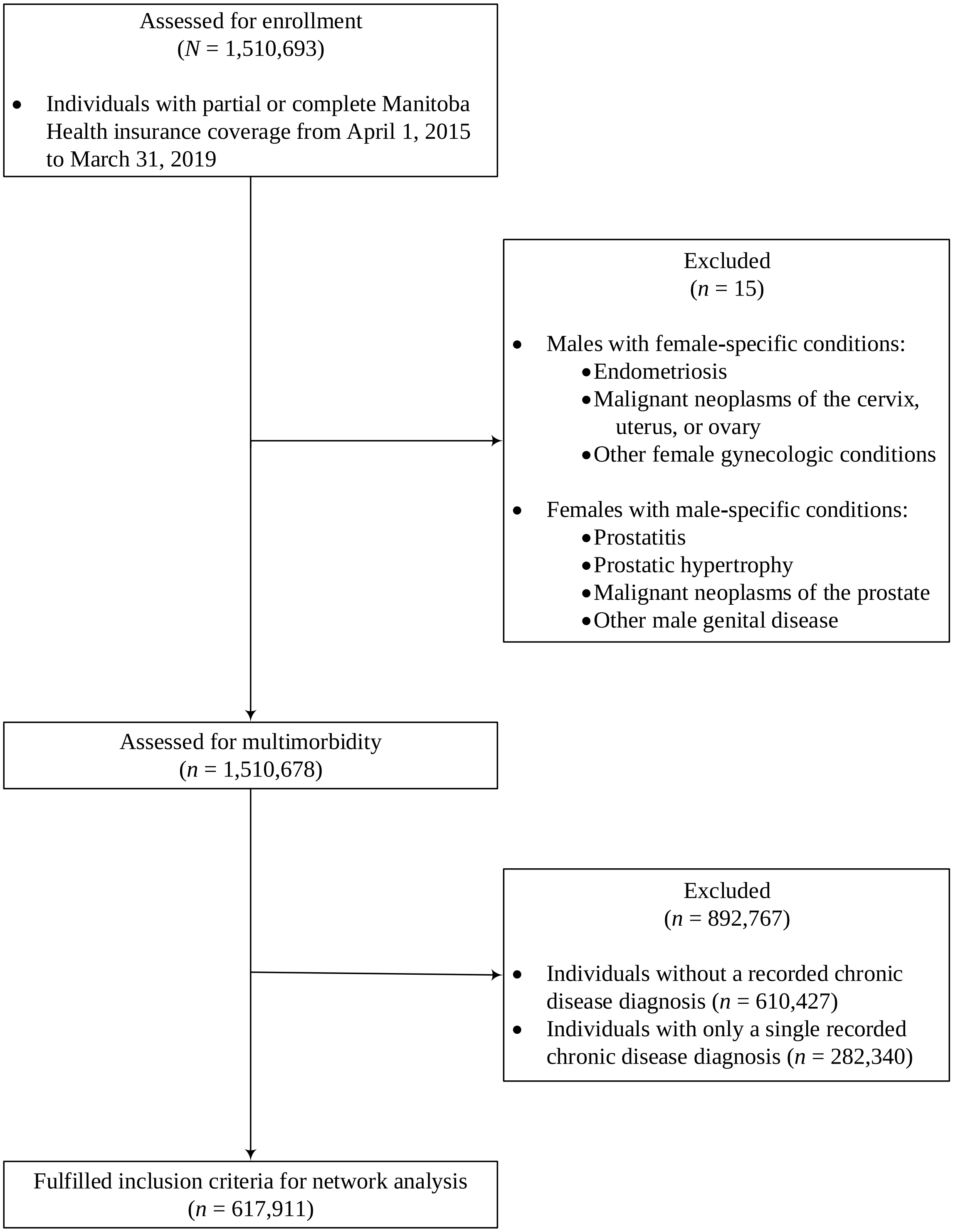
Sex	
Male	283,674 (45.9)
Female	334,237 (54.1)
Age (years)	
<20	43,072 (7.0)
20-39	102,750 (16.6)
49-59	189,300 (30.6)
60+	282,789 (45.8)
Residence locality	
Rural	221,923 (35.9)
Urban	395,907 (64.1)
Unknown	81 (<0.1)
Income quintile	
Q1 (lowest)	120,654 (19.5)
Q2	121,899 (19.7)
Q3	127,697 (20.7)
Q4	119,901 (19.4)
Q5 (highest)	115,384 (18.7)
Unknown	12,376 (2.0)
Healthcare utilization	
Inpatient hospitalization	83,934 (13.6)
Ambulatory visits	6 (3-10)
Chronic conditions	
2-3	304,084 (49.2)
4-5	150,938 (24.4)
6+	162,889 (26.4)

Data are presented as N (%) or median (Q1-Q3).

Demographic and chronic disease characteristics were measured at exit date.

Healthcare utilization was measured during the last 12 months of follow-up.

Figure 2. Participant flow diagram indicating the number of individuals excluded from the current study with explanation.



Assessed for enrollment
(*N* = 1,510,693)
• Individuals with partial or complete Manitoba Health insurance coverage from April 1, 2015 to March 31, 2019
Excluded
(*n* = 15)
• Males with female-specific conditions:
• Endometriosis
• Malignant neoplasms of the cervix, uterus, or ovary
• Other female gynecologic conditions
• Females with male-specific conditions:
• Prostatitis
• Prostatic hypertrophy
• Malignant neoplasms of the prostate
• Other male genital disease
Assessed for multimorbidity
(*n* = 1,510,678)
Excluded
(*n* = 892,767)
• Individuals without a recorded chronic disease diagnosis (*n* = 610,427)
• Individuals with only a single recorded chronic disease diagnosis (*n* = 282,340)
Fulfilled inclusion criteria for network analysis
(*n* = 617,911)

Note: Chronic disease ascertainment was performed using the Johns Hopkins ACG System.

4.1.2 Chronic Disease Characteristics

The five most prevalent MEDC categories were cardiovascular (29.1%), psychosocial (17.0%), endocrine (17.0%), musculoskeletal (12.7%), and allergy (9.4%). Hypertension was the most prevalent EDC category (22.5%) (Appendix C), followed by depression (11.1%), disorders of lipid metabolism (9.8%), degenerative joint disease (9.1%), type 2 diabetes mellitus (9.0%), and asthma (9.0%). Hypertension was the most prevalent EDC category among both males (22.2%) and females (22.9%) (Table 4). Following hypertension, the most prevalent EDC categories among males were disorders of lipid metabolism (10.5%), type 2 diabetes (9.4%), asthma (8.2%), depression (7.7%), and degenerative joint disease (7.5%); while depression (14.4%), degenerative joint disease (10.7%), asthma (9.9%), disorders of lipid metabolism (9.1%), and hypothyroidism (8.9%) were the next most prevalent conditions among females (Appendix D).

When the MEDC analyses were stratified by sex (Table 4), males had higher prevalence of genito-urinary (4.8% vs. 2.3%) and respiratory (8.1% vs. 7.5%) disorders; while females had higher prevalence in several MEDC categories including allergies (10.2% vs. 8.5%), endocrine disorders (20.5% vs. 13.6%), psychosocial disorders (20.2% vs. 13.9%), neurologic disorders (9.1% vs. 8.3%), musculoskeletal disorders (14.3% vs. 11.1%), gastrointestinal and hepatic disorders (8.1% vs. 6.4%), and hematologic disorders (4.6% vs. 3.0%). Compared with males, females had 7.2 times the amount of osteoporosis diagnoses, 3.1 times the amount of hypothyroidism diagnoses, and 2.5 times the number of rheumatoid arthritis diagnoses (Appendix D). Males had a significantly larger number of diagnoses for cardiomyopathy (80% higher), aortic aneurysm (60% higher), ischemic heart disease (50% higher), and acute myocardial infarction (50% higher).

Table 4. Frequency and prevalence of Major Expanded Diagnosis Clusters, ascertained using the Johns Hopkins ACG System, stratified by sex.

Major Expanded Diagnosis Cluster	Male (*n*=756,198)	Female (*n*=754,480)
Allergy	64,376 (8.5)	76,943 (10.2)
Cardiovascular	219,840 (29.1)	219,047 (29.0)
Dental	259 (0.0)	421 (0.1)
Ear, Nose, Throat	27,134 (3.6)	32,707 (4.3)
Endocrine	102,885 (13.6)	154,348 (20.5)
Eye	56,998 (7.5)	75,213 (10.0)
Female Reproductive	0 (0.0)	9,791 (1.3)
Gastrointestinal/Hepatic	48,297 (6.4)	61,394 (8.1)
General Signs and Symptoms	3,090 (0.4)	5,417 (0.7)
General Surgery	38,088 (5.0)	57,654 (7.6)
Genetic	24,304 (3.2)	22,615 (3.0)
Genito-urinary	36,006 (4.8)	17,472 (2.3)
Hematologic	22,471 (3.0)	34,778 (4.6)
Infections	1,408 (0.2)	998 (0.1)
Malignancies	32,778 (4.3)	34,266 (4.5)
Musculoskeletal	83,620 (11.1)	108,246 (14.3)
Neurologic	62,605 (8.3)	68,668 (9.1)
Nutrition	36,965 (4.9)	48,901 (6.5)
Psychosocial	105,299 (13.9)	152,086 (20.2)
Reconstructive	8,532 (1.1)	8,034 (1.1)
Renal	17,848 (2.4)	15,833 (2.1)
Respiratory	61,071 (8.1)	56,713 (7.5)
Rheumatologic	35,052 (4.6)	32,984 (4.4)
Skin	9,782 (1.3)	10,676 (1.4)

Data are presented as N (%).

Table 5. Number of disease co-occurrences identified, before and after statistically non-significant associations and negative correlations were excluded.

Disease ascertainment method	**Total before exclusions**	**Non-significant**	**Negative correlations**	**Total included**	**Pairwise associations**	**Triad associations**
Johns Hopkins EDC categorization	118,124	2,930	410	114,784	7,845	106,939
Elixhauser comorbidity index	4,407	28	4	4,375	449	3,926

Note: EDC = Expanded Diagnostic Cluster.

4.2 Disease Association Analysis

A total of 114,784 disease co-occurrences were identified using the Johns Hopkins ACG System, after non-significant (i.e., p-value > 0.01) and non-positive (i.e., phi < 0) associations were excluded (2.8%, *n*=3,340) (Table 5). Using the Elixhauser comorbidity index, 4,407 co-occurrences were identified after 0.7% (*n*=32) of associations were excluded (non-significant or non-positive). Hypergraphs were constructed using all 114,784 associations (both pairwise co-occurrences and triad associations). Pairwise disease networks were formed using 6.8% (*n*=7,845) and 10.3% (*n*=449) of all co-occurrences measured using the ACG System and Elixhauser index, respectively.

4.3 Global Network Properties

Since network density is not affected by edge weight, network density was constant (0.57) for all seven networks constructed with different co-occurrence measures when all edges (*n*=7,845) were included (Johns Hopkins ACG System, N nodes = 166). Similarly, network density was constant at 0.97 for the seven networks constructed based on the Elixhauser index (N nodes = 31). The smaller number of nodes included in the Elixhauser network, combined with a smaller number of low prevalent conditions (Appendix B), contributed to the Elixhauser-based network being significantly denser than the network based on the Johns Hopkins ACG System.

Networks constructed by limiting the number of associations using effect size cut-offs differed in network density and number of nodes (Table 6, Table 7). For pairwise networks constructed using the strongest 200 associations, the network with the least number of nodes (*n*=56, joint prevalence) had the highest network density (0.13), while the two networks with the greatest number of nodes (*n*=114, relative risk*;* *n*=123, Kulczynski) had the lowest network density at 0.03 (Table 6). As more associations were included, variation in the number of nodes and network density decreased between the networks. For the pairwise networks constructed with the strongest 50 percent of associations (*n*=3,922), the number of nodes ranged from 150 to 166 and network density varied between 0.29 and 0.35.

Among hypergraphs constructed using a defined number of hyperedges, the number of nodes and the percentage of triad associations varied (Table 7). In hypergraphs constructed from the top 200 associations, number of nodes ranged from 47 to 109 and the percentage of triad associations ranged from 28.5% to 99.0%. The hypergraph constructed using relative risk contained the smallest percentage of hyperedges relating to disease dyads (1.0%) and also contained the largest number of nodes (*n*=109); while the hypergraph based on joint prevalence had the smallest number of nodes (*n*=47) and also contained the largest percentage of dyad associations (71.5%). Hypergraphs constructed from the top 50 percent of associations had triad percentages ranging from 86.9% to 98.3% and had between 163 and 165 nodes.

Table 6. Global properties for pairwise networks constructed with select co-occurrence measures and limited to the strongest 200 statistically significant associations and the strongest 50 percent (*n*=3,922) of all statistically significant associations.

Association measure	**Top 200 associations**			**Top 50 percent of associations**		
	N nodes	N edges	Density	N nodes	N edges	Density
Lift	108	200	0.04	165	3,922	0.29
Relative risk	114	200	0.03	166	3,922	0.29
Phi	87	200	0.05	164	3,922	0.29
Jaccard	72	200	0.08	150	3,922	0.35
Cosine	73	200	0.08	161	3,922	0.31
Kulczynski	123	200	0.03	166	3,922	0.29
Joint prevalence	56	200	0.13	151	3,922	0.35

Note: Chronic diseases where ascertained using the Johns Hopkins ACG System.

Table 7. Global properties for chronic disease hypergraphs constructed with select co-occurrence measures and limited to the strongest 200 statistically significant associations and the strongest 50 percent (*n*=3,922) of all statistically significant associations.

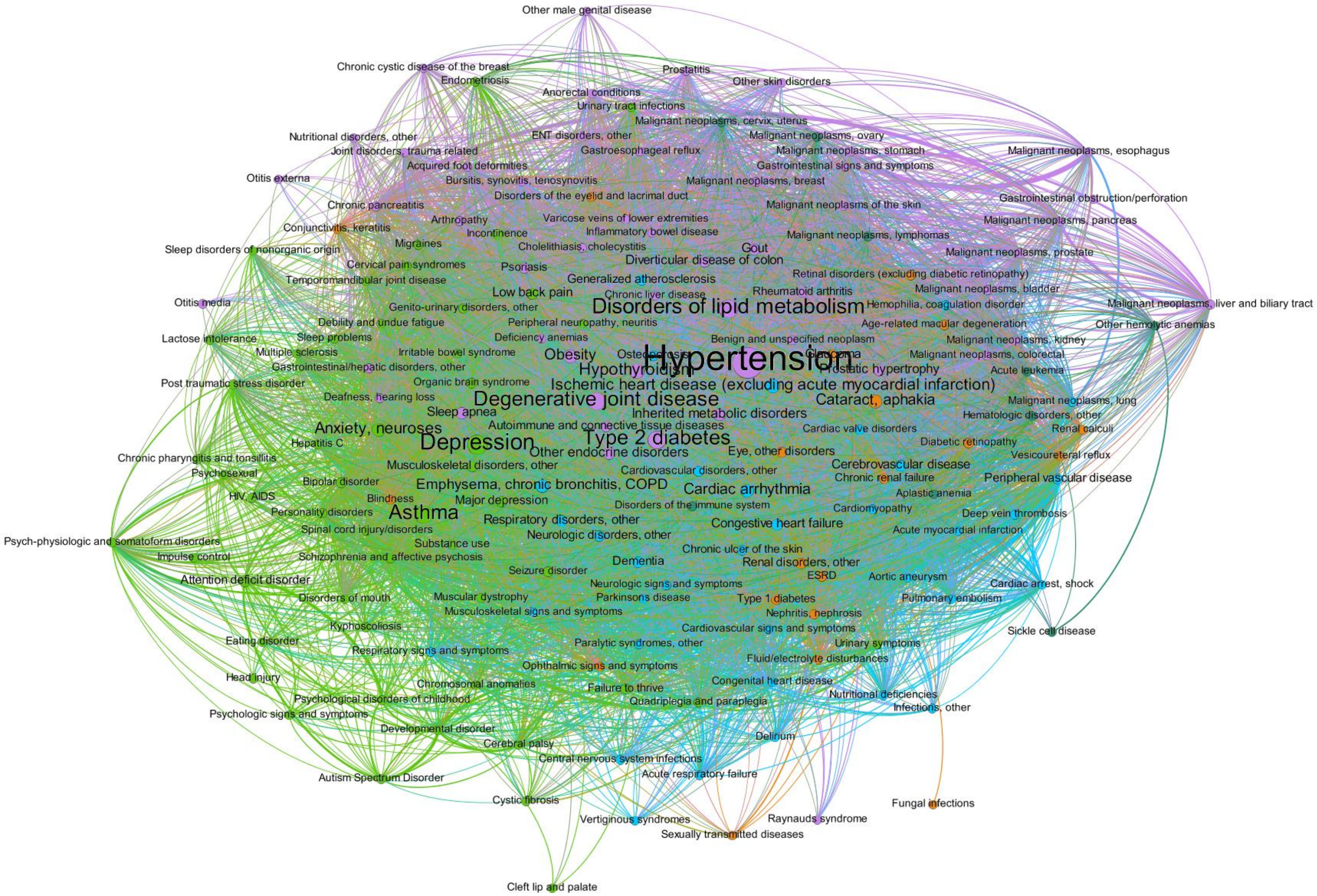
Association measure	**Top 200 associations**			**Top 50 percent of associations**		
	N nodes	N hyperedges	N (%) triads	N nodes	N hyperedges	N (%) triads
Lift	97	200	195 (97.5)	165	57,392	55,597 (96.9)
Relative risk	109	200	198 (99.0)	165	57,392	56,060 (97.7)
Phi	55	200	125 (62.5)	164	57,392	53,157 (92.6)
Jaccard	65	200	57 (28.5)	163	57,392	49,891 (86.9)
Cosine	53	200	107 (53.5)	164	57,392	52,284 (91.1)
Kulczynski	107	200	195 (97.5)	165	57,392	56,440 (98.3)
Joint prevalence	47	200	57 (28.5)	163	57,392	51,273 (89.3)

4.4 Network Visualization

Including all statistically significant pairwise associations for the 167 disease categories obtained using the Johns Hopkins ACG System (Figure 3) and the 31 Elixhauser comorbidities (Figure 4), produced dense network visualizations that are difficult to interpret. Reducing complexity by selecting the strongest (i.e., highest effect size) 200 EDC associations produced more interpretable network diagrams (Figure 5, Figure 6, Figure 7, Figure 8, Figure 9, Figure 10, Figure 11). Visual interpretability of the disease networks limited to the top 200 co-occurrence relationships varied depending on the association measure used to construct the network.

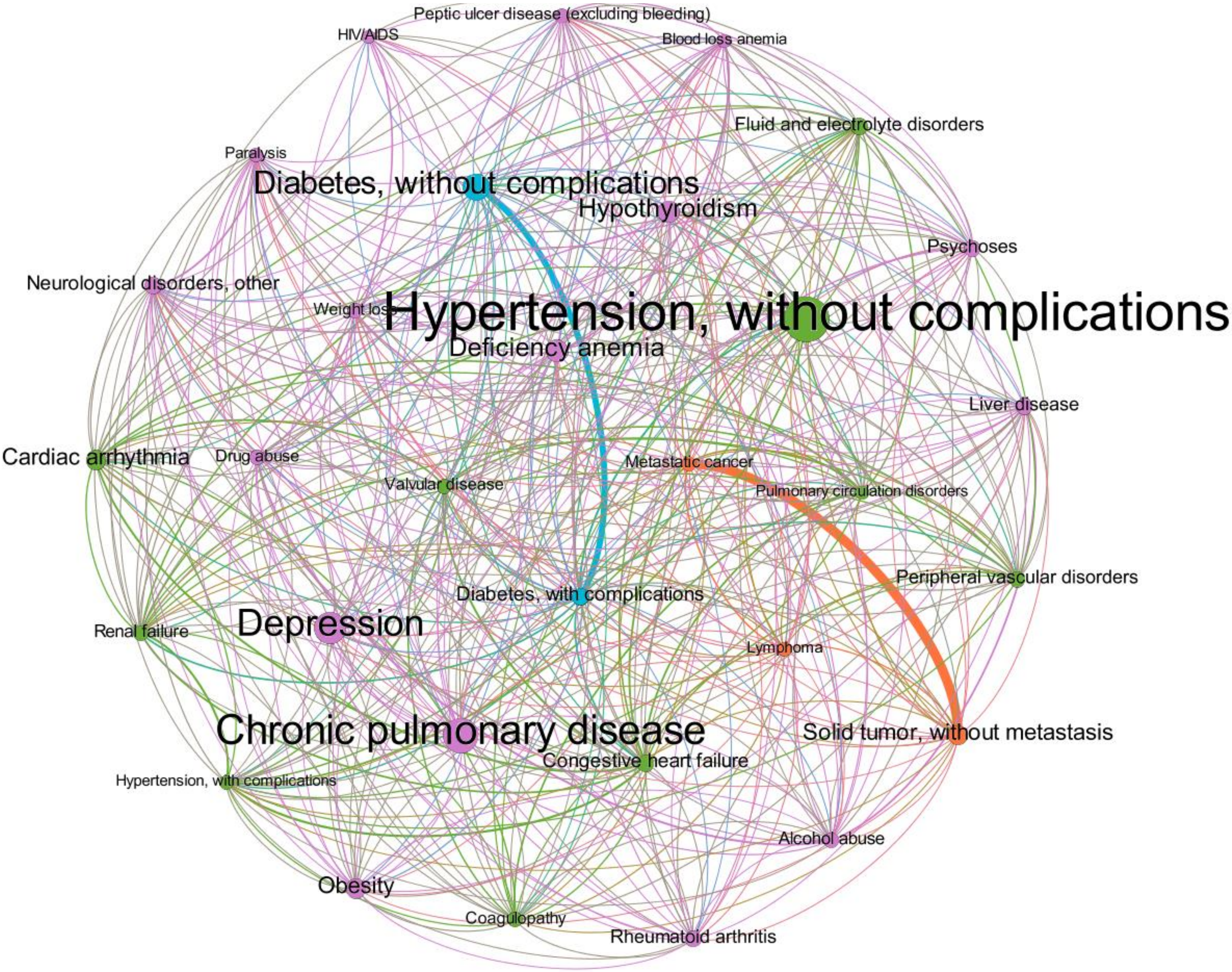
The traditional hypergraph visualization (Figure 12), which uses coloured bounding containers to represent hyperedges, has reduced visual interpretability compared with pairwise graphs even with a reduced number of visualized co-occurrence relationships (*n*=30). Visual interpretability of hypergraphs is reduced due to increased network complexity, overlapping elements, and the need to reuse colours for multiple hyperedges. Compared with pairwise graphs, the increased complexity of hypergraph figures makes it more difficult to visualize edge weights and incorporate inline node labels. The PAOH visualization (Figure 13), an alternative hypergraph visualization which uses vertical bars to represent hyperedges, allows for the incorporation of a higher number of relationships (*n*=100) and produces more visually discernable patterns, but currently does not provide support for edge weights.

Figure 3. Pairwise chronic disease network constructed using all statistically significant associations (*n*=7,845), with diseases ascertained using the Johns Hopkins ACG System and co-occurrence relationships measured using phi.


Other male genital disease
Chronic cystic disease of the breast
Endometriosis
Prostatitis
Other skin disorders
Anorectal conditions
Urinary tract infections
Malignant neoplasms, cervix, uterus
Nutritional disorders, other
ENT disorders, other
Malignant neoplasms, ovary
Joint disorders, trauma related
Gastroesophageal reflux
Malignant neoplasms, stomach
Malignant neoplasms, esophagus
Acquired foot deformities
Gastrointestinal signs and symptoms
Otitis externa
Bursitis, synovitis, tenosynovitis
Malignant neoplasms, breast
Disorders of the eyelid and lacrimal duct
Gastrointestinal obstruction/perforation
Chronic pancreatitis
Malignant neoplasms of the skin
Arthropathy
Varicose veins of lower extremities
Malignant neoplasms, pancreas
Conjunctivitis, keratitis
Incontinence
Inflammatory bowel disease
Malignant neoplasms, lymphomas
Migraines
Cholelithiasis, cholecystitis
Gout
Malignant neoplasms, prostate
Sleep disorders of nonorganic origin
Diverticular disease of colon
Cervical pain syndromes
Psoriasis
Retinal disorders (excluding diabetic retinopathy)
Temporomandibular joint disease
Generalized atherosclerosis
Malignant neoplasms, bladder
Low back pain
Chronic liver disease
Rheumatoid arthritis
Otitis media
Malignant neoplasms, liver and biliary tract
Genito-urinary disorders, other
Disorders of lipid metabolism
Hemophilia, coagulation disorder
Debility and undue fatigue
Peripheral neuropathy, neuritis
Age-related macular degeneration
Other hemolytic anemias
Lactose intolerance
Sleep problems
Deficiency anemias
Benign and unspecified neoplasm
Malignant neoplasms, kidney
Multiple sclerosis
Irritable bowel syndrome
Obesity
Osteoporosis
Hypertension
Glaucoma
Malignant neoplasms, colorectal
Gastrointestinal/hepatic disorders, other
Hypothyroidism
Prostatic hypertrophy
Acute leukemia
Post traumatic stress disorder
Organic brain syndrome
Ischemic heart disease (excluding acute myocardial infarction)
Deafness, hearing loss
Degenerative joint disease
Cataract, aphakia
Malignant neoplasms, lung
Sleep apnea
Inherited metabolic disorders
Hematologic disorders, other
Anxiety, neuroses
Autoimmune and connective tissue diseases
Cardiac valve disorders
Renal calculi
Hepatitis C
Depression
Type 2 diabetes
Diabetic retinopathy
Other endocrine disorders
Eye, other disorders
Chronic pharyngitis and tonsillitis
Vesicoureteral reflux
Musculoskeletal disorders, other
Cerebrovascular disease
Psychosexual
Cardiovascular disorders, other
Chronic renal failure
Peripheral vascular disease
Bipolar disorder
Emphysema, chronic bronchitis, COPD
Cardiac arrhythmia
HIV, AIDS
Aplastic anemia
Blindness
Major depression
Disorders of the immune system
Cardiomyopathy
Personality disorders
Asthma
Deep vein thrombosis
Respiratory disorders, other
Congestive heart failure
Spinal cord injury/disorders
Acute myocardial infarction
Neurologic disorders, other
Psych-physiologic and somatoform disorders
Substance use
Chronic ulcer of the skin
Impulse control
Schizophrenia and affective psychosis
Dementia
Renal disorders, other
Seizure disorder
ESRD
Aortic aneurysm
Attention deficit disorder
Neurologic signs and symptoms
Cardiac arrest, shock
Disorders of mouth
Muscular dystrophy
Parkinsons disease
Type 1 diabetes
Pulmonary embolism
Musculoskeletal signs and symptoms
Nephritis, nephrosis
Kyphoscoliosis
Cardiovascular signs and symptoms
Sickle cell disease
Eating disorder
Paralytic syndromes, other
Urinary symptoms
Respiratory signs and symptoms
Fluid/electrolyte disturbances
Ophthalmic signs and symptoms
Head injury
Congenital heart disease
Chromosomal anomalies
Failure to thrive
Nutritional deficiencies
Psychological disorders of childhood
Quadriplegia and paraplegia
Infections, other
Psychologic signs and symptoms
Developmental disorder
Delirium
Cerebral palsy
Central nervous system infections
Acute respiratory failure
Autism Spectrum Disorder
Cystic fibrosis
Fungal infections
Vertiginous syndromes
Raynauds syndrome
Sexually transmitted diseases
Cleft lip and palate

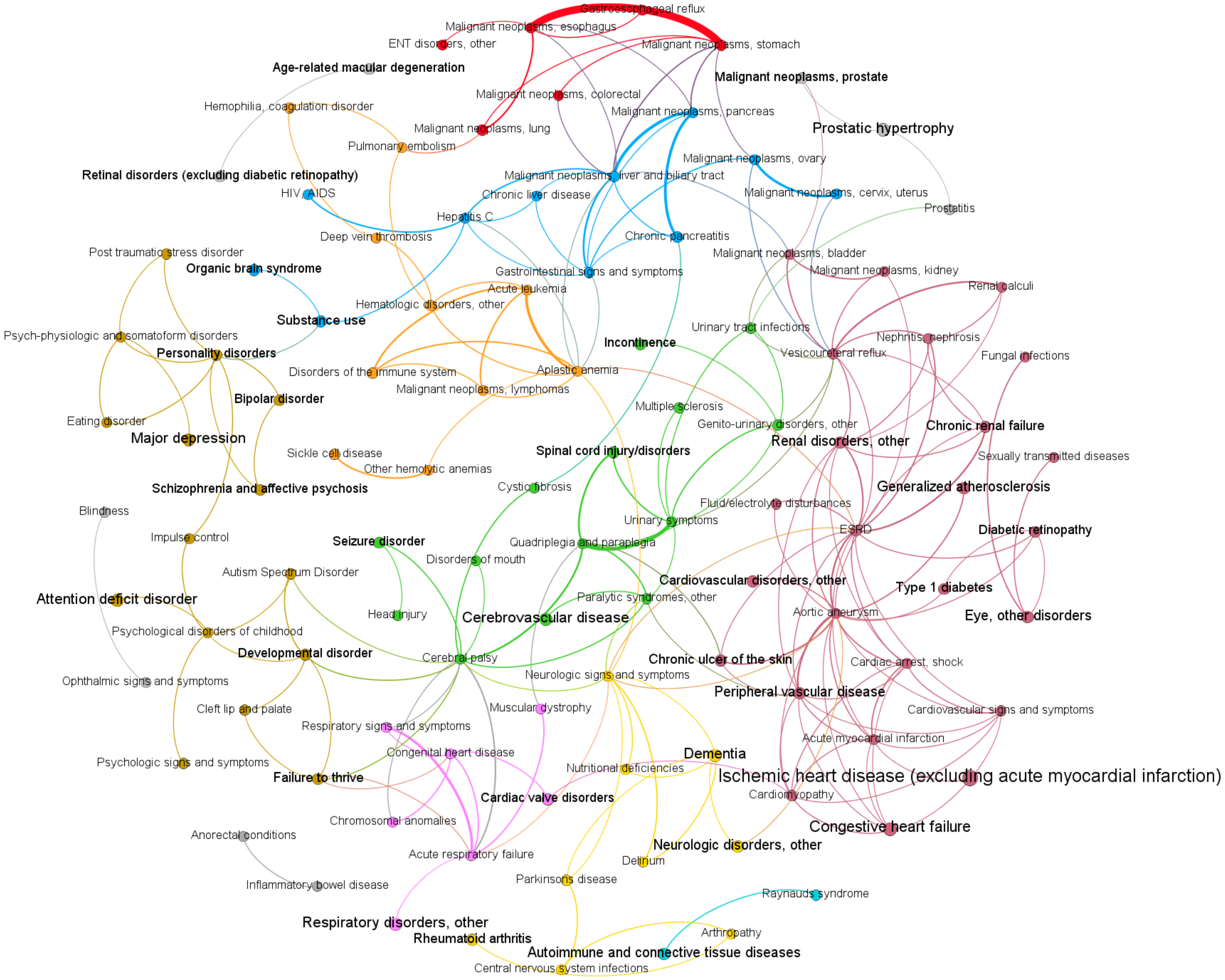
Note: Node diameter and font size are proportional to disease prevalence, edge weight (thickness) is proportional to effect size, and node and edge colour indicate community structure (i.e., disease clusters). COPD = chronic obstructive pulmonary disease; ENT = ear, nose, and throat; ESRD = end-stage renal disease; HIV/AIDS = human immunodeficiency virus/acquired immunodeficiency syndrome.

Figure 4. Pairwise chronic disease network constructed using all statistically significant associations (*n*=449), with diseases ascertained using the Elixhauser comorbidity index and co-occurrence relationships measured using phi.


Peptic ulcer disease (excluding bleeding)
HIV/AIDS
Blood loss anemia
Fluid and electrolyte disorders
Paralysis
Diabetes, without complications
Hypothyroidism
Psychoses
Neurological disorders, other
Weight loss
Hypertension, without complications
Deficiency anemia
Liver disease
Cardiac arrhythmia
Drug abuse
Metastatic cancer
Valvular disease
Pulmonary circulation disorders
Peripheral vascular disorders
Diabetes, with complications
Renal failure
Depression
Lymphoma
Chronic pulmonary disease
Solid tumor, without metastasis
Congestive heart failure
Hypertension, with complications
Alcohol abuse
Obesity
Coagulopathy
Rheumatoid arthritis

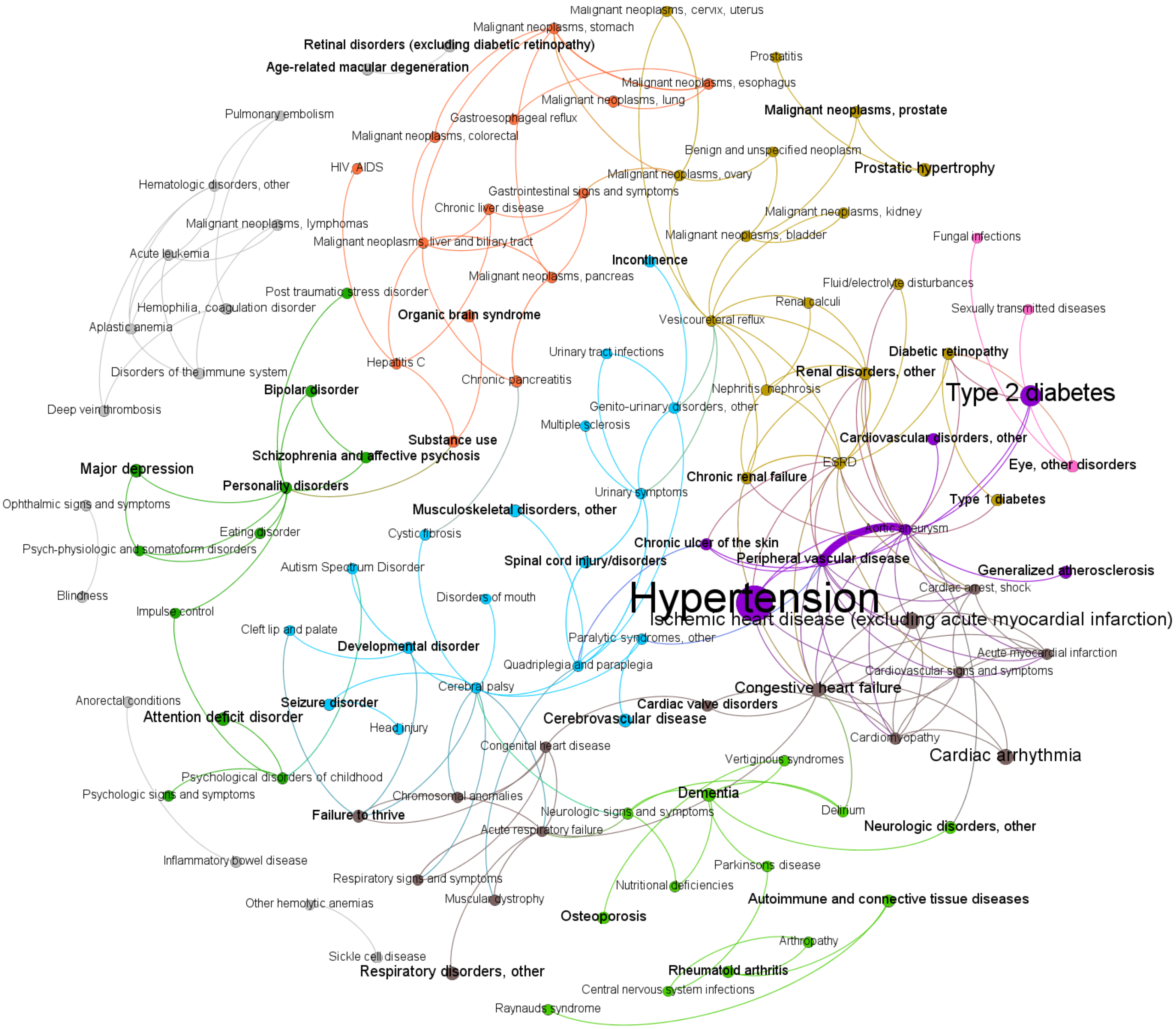
Note: Node diameter and font size are proportional to disease prevalence, edge weight (thickness) is proportional to effect size, and node and edge colour indicate community structure (i.e., disease clusters). HIV/AIDS = human immunodeficiency virus/acquired immunodeficiency syndrome.

Figure 5. Pairwise chronic disease network with co-occurrence relationships measured using lift and diseases ascertained using the Johns Hopkins ACG System, limited to the strongest 200 statistically significant associations.


Gastroesophageal reflux
Malignant neoplasms, esophagus
ENT disorders, other
Malignant neoplasms, stomach
Age-related macular degeneration
Malignant neoplasms, prostate
Malignant neoplasms, colorectal
Hemophilia, coagulation disorder
Malignant neoplasms, pancreas
Malignant neoplasms, lung
Prostatic hypertrophy
Pulmonary embolism
Malignant neoplasms, ovary
Retinal disorders (excluding diabetic retinopathy)
Malignant neoplasms, liver and biliary tract
HIV/AIDS
Chronic liver disease
Malignant neoplasms, cervix, uterus
Hepatitis C
Prostatitis
Deep vein thrombosis
Chronic pancreatitis
Post traumatic stress disorder
Malignant neoplasms, bladder
Organic brain syndrome
Gastrointestinal signs and symptoms
Malignant neoplasms, kidney
Acute leukemia
Renal calculi
Hematologic disorders, other
Substance use
Urinary tract infections
Psych-physiologic and somatoform disorders
Incontinence
Nephritis, nephrosis
Personality disorders
Vesicoureteral reflux
Fungal infections
Disorders of the immune system
Aplastic anemia
Malignant neoplasms, lymphomas
Bipolar disorder
Multiple sclerosis
Eating disorder
Genito-urinary disorders, other
Chronic renal failure
Major depression
Renal disorders, other
Spinal cord injury/disorders
Sickle cell disease
Sexually transmitted diseases
Other hemolytic anemias
Schizophrenia and affective psychosis
Cystic fibrosis
Generalized atherosclerosis
Fluid/electrolyte disturbances
Blindness
Urinary symptoms
ESRD
Diabetic retinopathy
Impulse control
Seizure disorder
Quadriplegia and paraplegia
Disorders of mouth
Autism Spectrum Disorder
Cardiovascular disorders, other
Type 1 diabetes
Attention deficit disorder
Paralytic syndromes, other
Head injury
Cerebrovascular disease
Aortic aneurysm
Eye, other disorders
Psychological disorders of childhood
Developmental disorder
Cerebral palsy
Chronic ulcer of the skin
Cardiac arrest, shock
Ophthalmic signs and symptoms
Neurologic signs and symptoms
Peripheral vascular disease
Cleft lip and palate
Muscular dystrophy
Cardiovascular signs and symptoms
Respiratory signs and symptoms
Acute myocardial infarction
Congenital heart disease
Dementia
Psychologic signs and symptoms
Nutritional deficiencies
Ischemic heart disease (excluding acute myocardial infarction)
Failure to thrive
Cardiac valve disorders
Cardiomyopathy
Chromosomal anomalies
Congestive heart failure
Anorectal conditions
Neurologic disorders, other
Acute respiratory failure
Delirium
Parkinsons disease
Inflammatory bowel disease
Raynauds syndrome
Respiratory disorders, other
Arthropathy
Rheumatoid arthritis
Autoimmune and connective tissue diseases
Central nervous system infections

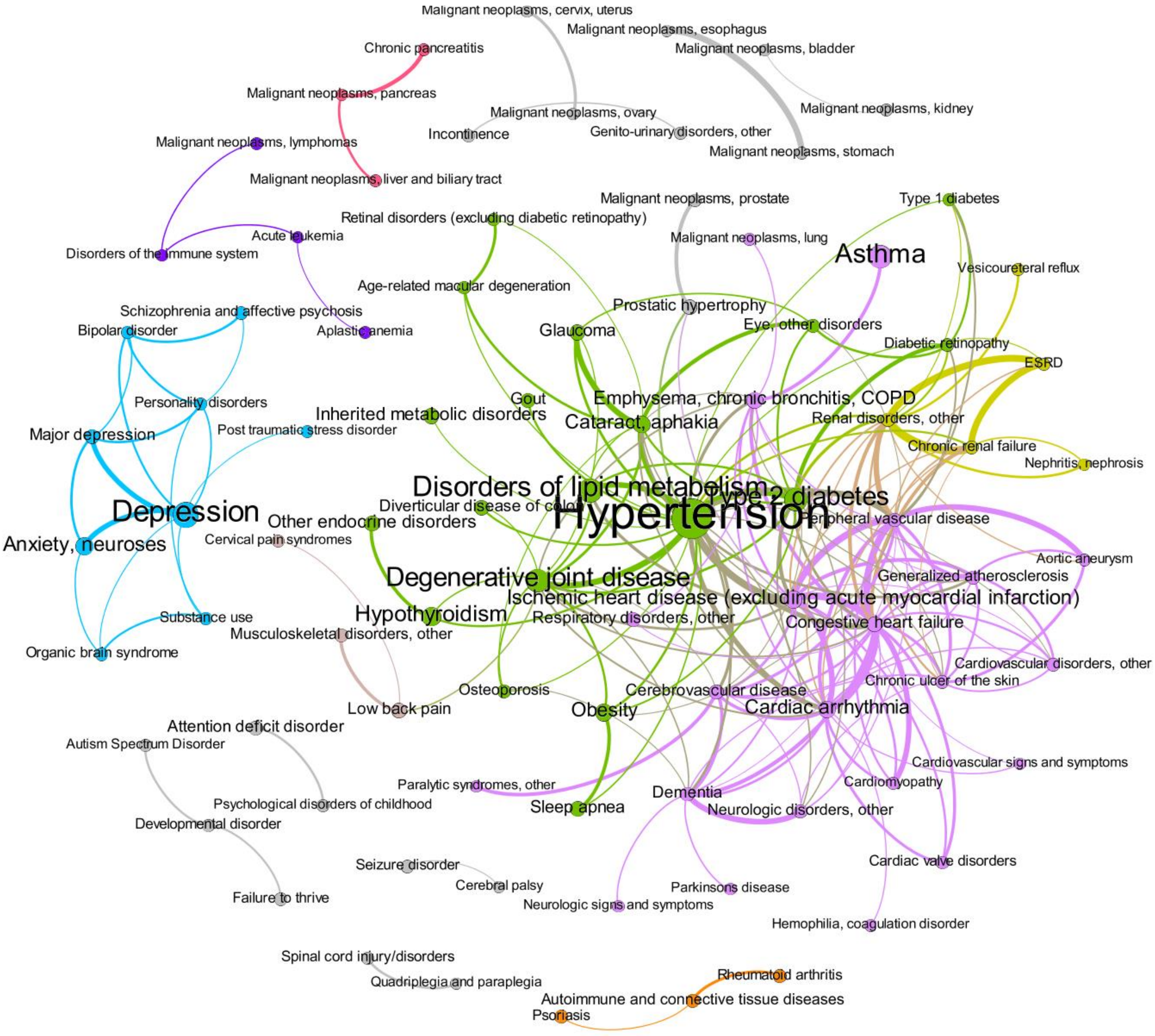
Note: Node diameter and font size are proportional to disease prevalence, edge weight (thickness) is proportional to effect size, and node and edge colour indicate community structure (i.e., disease clusters). ENT = ear, nose, and throat; ESRD = end-stage renal disease; HIV/AIDS = human immunodeficiency virus/acquired immunodeficiency syndrome.

Figure 6. Pairwise chronic disease network with co-occurrence relationships measured using relative risk and diseases ascertained using the Johns Hopkins ACG System, limited to the strongest 200 statistically significant associations.


Malignant neoplasms, cervix, uterus
Malignant neoplasms, stomach
Retinal disorders (excluding diabetic retinopathy)
Prostatitis
Age-related macular degeneration
Malignant neoplasms, esophagus
Malignant neoplasms, lung
Malignant neoplasms, prostate
Pulmonary embolism
Gastroesophageal reflux
Malignant neoplasms, colorectal
Benign and unspecified neoplasm
HIV/AIDS
Prostatic hypertrophy
Malignant neoplasms, ovary
Hematologic disorders, other
Gastrointestinal signs and symptoms
Chronic liver disease
Malignant neoplasms, kidney
Malignant neoplasms, lymphomas
Malignant neoplasms, bladder
Fungal infections
Malignant neoplasms, liver and biliary tract
Acute leukemia
Incontinence
Malignant neoplasms, pancreas
Fluid/electrolyte disturbances
Post traumatic stress disorder
Renal calculi
Sexually transmitted diseases
Hemophilia, coagulation disorder
Organic brain syndrome
Aplastic anemia
Vesicoureteral reflux
Urinary tract infections
Diabetic retinopathy
Hepatitis C
Renal disorders, other
Disorders of the immune system
Chronic pancreatitis
Type 2 diabetes
Bipolar disorder
Nephritis, nephrosis
Genito-urinary disorders, other
Deep vein thrombosis
Multiple sclerosis
Substance use
Cardiovascular disorders, other
Schizophrenia and affective psychosis
ESRD
Eye, other disorders
Major depression
Chronic renal failure
Personality disorders
Urinary symptoms
Type 1 diabetes
Ophthalmic signs and symptoms
Musculoskeletal disorders, other
Aortic aneurysm
Eating disorder
Cystic fibrosis
Chronic ulcer of the skin
Psych-physiologic and somatoform disorders
Peripheral vascular disease
Spinal cord injury/disorders
Autism Spectrum Disorder
Generalized atherosclerosis
Cardiac arrest, shock
Blindness
Disorders of mouth
Hypertension
Impulse control
Ischemic heart disease (excluding acute myocardial infarction)
Cleft lip and palate
Paralytic syndromes, other
Developmental disorder
Acute myocardial infarction
Quadriplegia and paraplegia
Cardiovascular signs and symptoms
Cerebral palsy
Congestive heart failure
Anorectal conditions
Seizure disorder
Cardiac valve disorders
Attention deficit disorder
Cerebrovascular disease
Head injury
Cardiomyopathy
Congenital heart disease
Vertiginous syndromes
Cardiac arrhythmia
Psychological disorders of childhood
Dementia
Psychologic signs and symptoms
Chromosomal anomalies
Delirium
Neurologic signs and symptoms
Failure to thrive
Neurologic disorders, other
Acute respiratory failure
Inflammatory bowel disease
Parkinsons disease
Respiratory signs and symptoms
Nutritional deficiencies
Other hemolytic anemias
Muscular dystrophy
Autoimmune and connective tissue diseases
Osteoporosis
Arthropathy
Sickle cell disease
Rheumatoid arthritis
Respiratory disorders, other
Central nervous system infections
Raynauds syndrome

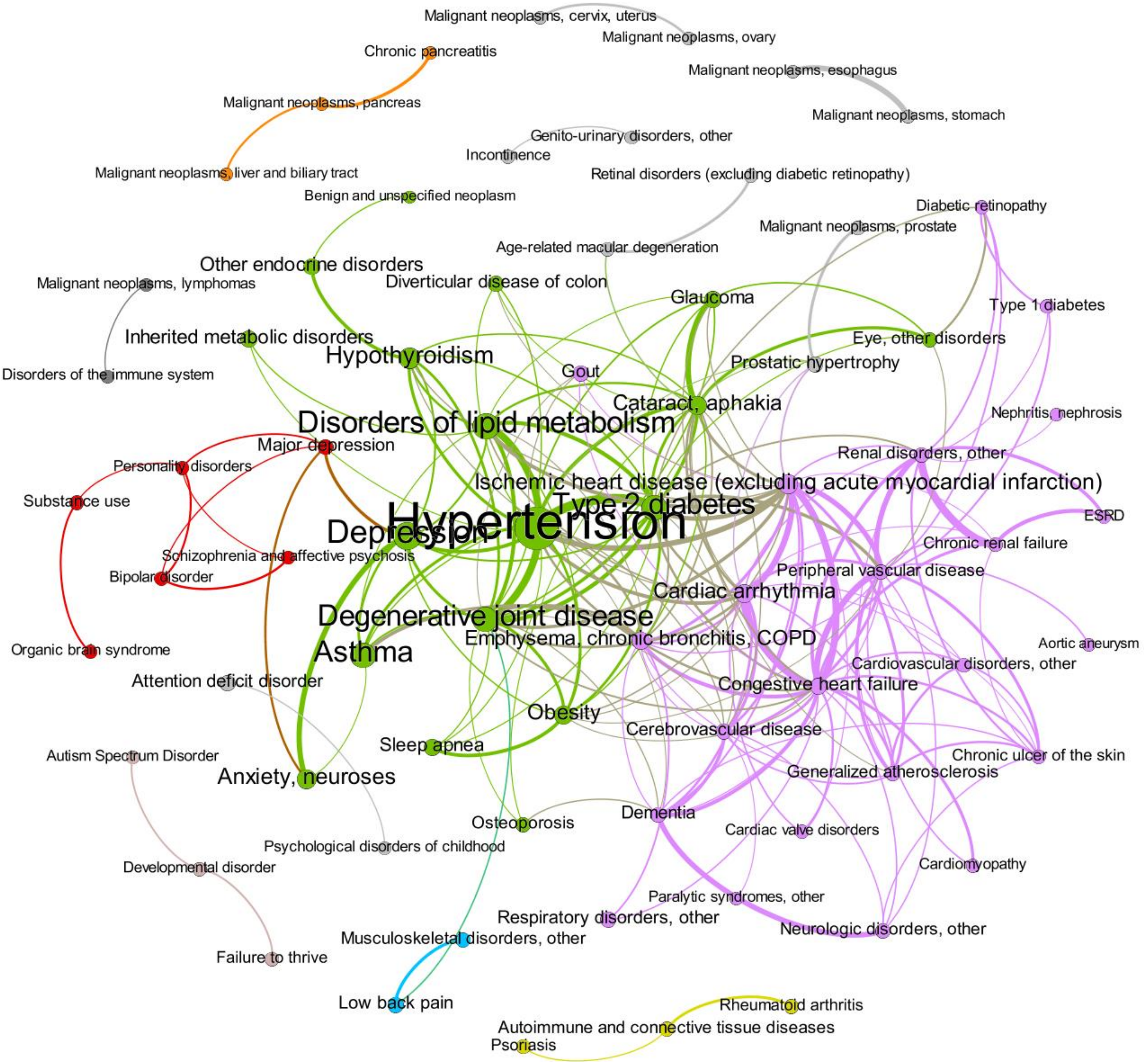
Note: Node diameter and font size are proportional to disease prevalence, edge weight (thickness) is proportional to effect size, and node and edge colour indicate community structure (i.e., disease clusters). ESRD = end-stage renal disease, HIV/AIDS = human immunodeficiency virus/acquired immunodeficiency syndrome.

Figure 7. Pairwise chronic disease network with co-occurrence relationships measured using phi and diseases ascertained using the Johns Hopkins ACG System, limited to the strongest 200 statistically significant associations.


Malignant neoplasms, cervix, uterus
Malignant neoplasms, esophagus
Chronic pancreatitis
Malignant neoplasms, bladder
Malignant neoplasms, pancreas
Malignant neoplasms, kidney
Malignant neoplasms, ovary
Incontinence
Genito-urinary disorders, other
Malignant neoplasms, lymphomas
Malignant neoplasms, stomach
Malignant neoplasms, liver and biliary tract
Malignant neoplasms, prostate
Type 1 diabetes
Retinal disorders (excluding diabetic retinopathy)
Acute leukemia
Malignant neoplasms, lung
Disorders of the immune system
Asthma
Vesicoureteral reflux
Age-related macular degeneration
Schizophrenia and affective psychosis
Prostatic hypertrophy
Bipolar disorder
Aplastic anemia
Glaucoma
Eye, other disorders
Diabetic retinopathy
ESRD
Personality disorders
Gout
Emphysema, chronic bronchitis, COPD
Inherited metabolic disorders
Renal disorders, other
Major depression
Post traumatic stress disorder
Cataract, aphakia
Chronic renal failure
Nephritis, nephrosis
Disorders of lipid metabolism
Type 2 diabetes
Diverticular disease of colon
Depression
Hypertension
Other endocrine disorders
Peripheral vascular disease
Anxiety, neuroses
Cervical pain syndromes
Aortic aneurysm
Degenerative joint disease
Generalized atherosclerosis
Ischemic heart disease (excluding acute myocardial infarction)
Hypothyroidism
Substance use
Respiratory disorders, other
Congestive heart failure
Musculoskeletal disorders, other
Organic brain syndrome
Cardiovascular disorders, other
Chronic ulcer of the skin
Osteoporosis
Cerebrovascular disease
Low back pain
Obesity
Cardiac arrhythmia
Attention deficit disorder
Autism Spectrum Disorder
Cardiovascular signs and symptoms
Paralytic syndromes, other
Cardiomyopathy
Dementia
Psychological disorders of childhood
Sleep apnea
Neurologic disorders, other
Developmental disorder
Seizure disorder
Cardiac valve disorders
Cerebral palsy
Parkinsons disease
Failure to thrive
Neurologic signs and symptoms
Hemophilia, coagulation disorder
Spinal cord injury/disorders
Rheumatoid arthritis
Quadriplegia and paraplegia
Autoimmune and connective tissue diseases
Psoriasis

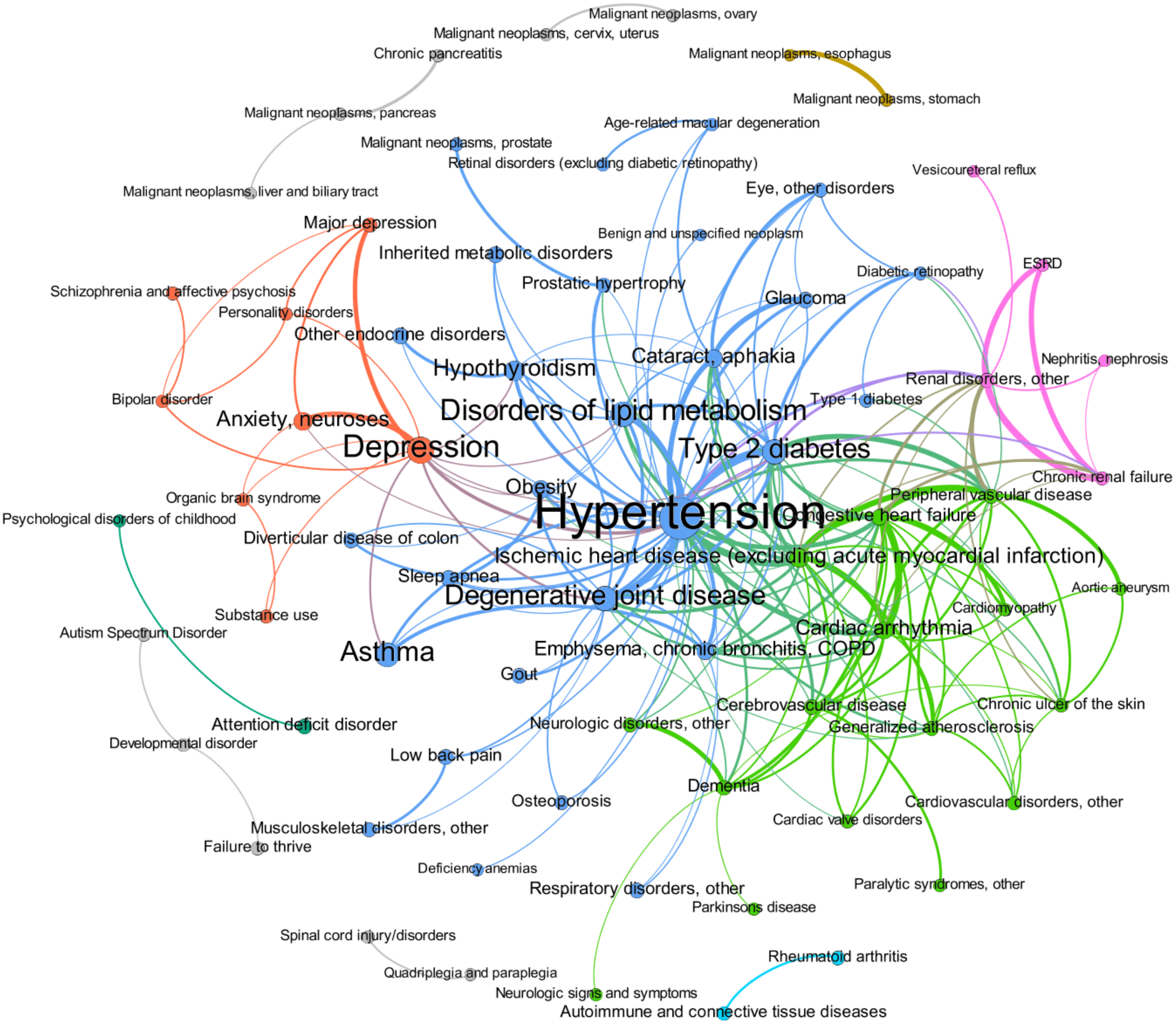
Note: Node diameter and font size are proportional to disease prevalence, edge weight (thickness) is proportional to effect size, and node and edge colour indicate community structure (i.e., disease clusters). COPD = chronic obstructive pulmonary disease, ESRD = end-stage renal disease.

Figure 8. Pairwise chronic disease network with co-occurrence relationships measured using Jaccard and diseases ascertained using the Johns Hopkins ACG System, limited to the strongest 200 statistically significant associations.


Malignant neoplasms, cervix, uterus
Malignant neoplasms, ovary
Chronic pancreatitis
Malignant neoplasms, esophagus
Malignant neoplasms, pancreas
Malignant neoplasms, stomach
Genito-urinary disorders, other
Incontinence
Malignant neoplasms, liver and biliary tract
Retinal disorders (excluding diabetic retinopathy)
Benign and unspecified neoplasm
Diabetic retinopathy
Malignant neoplasms, prostate
Age-related macular degeneration
Other endocrine disorders
Malignant neoplasms, lymphomas
Diverticular disease of colon
Glaucoma
Type 1 diabetes
Inherited metabolic disorders
Eye, other disorders
Hypothyroidism
Prostatic hypertrophy
Disorders of the immune system
Gout
Cataract, aphakia
Nephritis, nephrosis
Disorders of lipid metabolism
Major depression
Renal disorders, other
Personality disorders
Ischemic heart disease (excluding acute myocardial infarction)
Substance use
Type 2 diabetes
Hypertension
ESRD
Depression
Chronic renal failure
Schizophrenia and affective psychosis
Peripheral vascular disease
Bipolar disorder
Cardiac arrhythmia
Degenerative joint disease
Organic brain syndrome
Emphysema, chronic bronchitis, COPD
Aortic aneurysm
Asthma
Cardiovascular disorders, other
Attention deficit disorder
Congestive heart failure
Obesity
Cerebrovascular disease
Sleep apnea
Chronic ulcer of the skin
Autism Spectrum Disorder
Generalized atherosclerosis
Anxiety, neuroses
Dementia
Osteoporosis
Cardiac valve disorders
Psychological disorders of childhood
Developmental disorder
Cardiomyopathy
Paralytic syndromes, other
Respiratory disorders, other
Neurologic disorders, other
Musculoskeletal disorders, other
Failure to thrive
Low back pain
Rheumatoid arthritis
Autoimmune and connective tissue diseases
Psoriasis

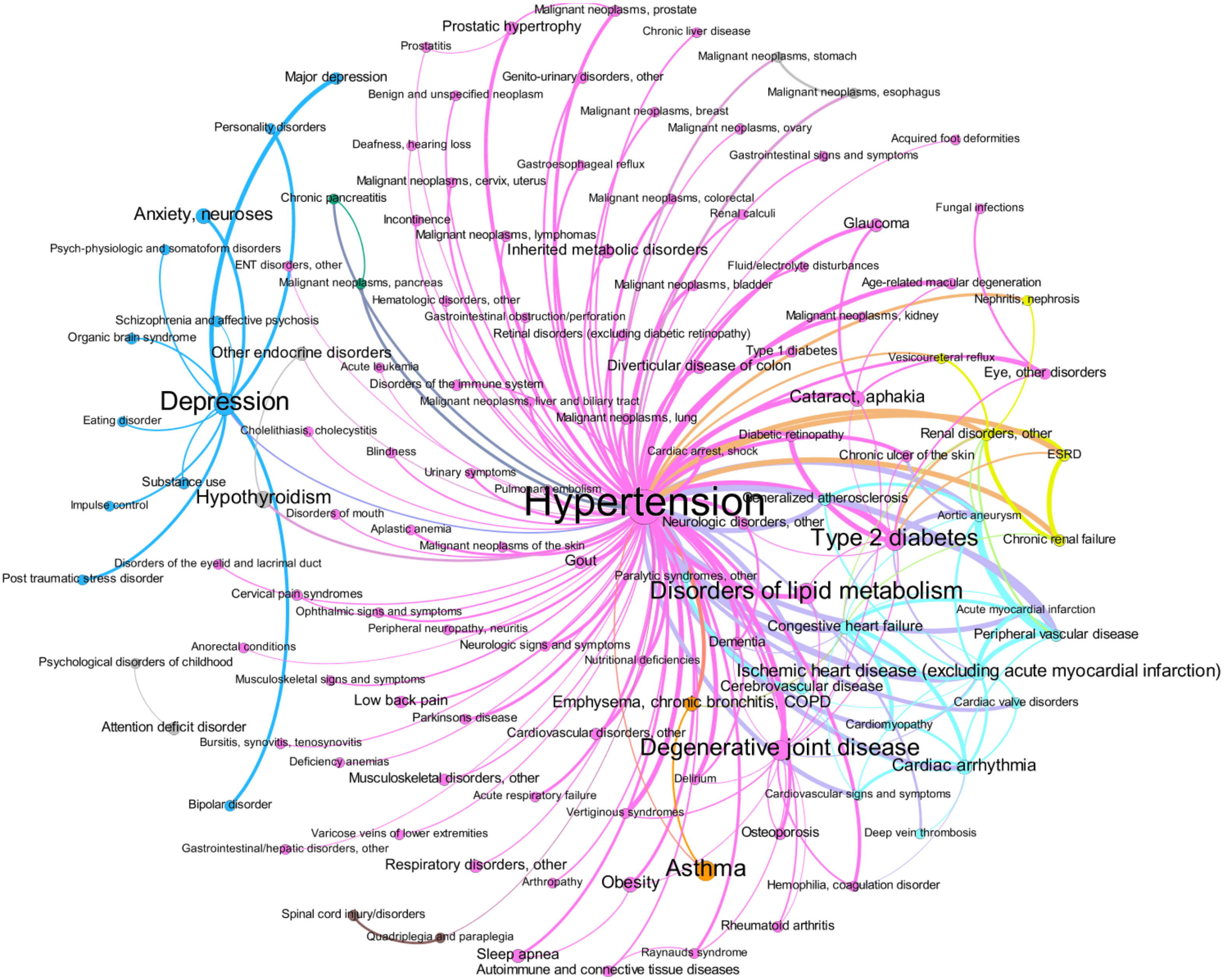
Note: Node diameter and font size are proportional to disease prevalence, edge weight (thickness) is proportional to effect size, and node and edge colour indicate community structure (i.e., disease clusters). COPD = chronic obstructive pulmonary disease, ESRD = end-stage renal disease.

Figure 9. Pairwise chronic disease network with co-occurrence relationships measured using cosine and diseases ascertained using the Johns Hopkins ACG System, limited to the strongest 200 statistically significant associations.


Malignant neoplasms, ovary
Malignant neoplasms, cervix, uterus
Chronic pancreatitis
Malignant neoplasms, esophagus
Malignant neoplasms, stomach
Malignant neoplasms, pancreas
Age-related macular degeneration
Malignant neoplasms, prostate
Retinal disorders (excluding diabetic retinopathy)
Vesicoureteral reflux
Malignant neoplasms, liver and biliary tract
Eye, other disorders
Major depression
Benign and unspecified neoplasm
Inherited metabolic disorders
Diabetic retinopathy
ESRD
Schizophrenia and affective psychosis
Prostatic hypertrophy
Glaucoma
Personality disorders
Other endocrine disorders
Cataract, aphakia
Nephritis, nephrosis
Hypothyroidism
Renal disorders, other
Bipolar disorder
Type 1 diabetes
Disorders of lipid metabolism
Anxiety, neuroses
Depression
Type 2 diabetes
Chronic renal failure
Organic brain syndrome
Obesity
Peripheral vascular disease
Psychological disorders of childhood
Hypertension
Congestive heart failure
Diverticular disease of colon
Ischemic heart disease (excluding acute myocardial infarction)
Sleep apnea
Aortic aneurysm
Degenerative joint disease
Cardiomyopathy
Substance use
Cardiac arrhythmia
Autism Spectrum Disorder
Asthma
Emphysema, chronic bronchitis, COPD
Gout
Cerebrovascular disease
Chronic ulcer of the skin
Attention deficit disorder
Neurologic disorders, other
Generalized atherosclerosis
Developmental disorder
Low back pain
Dementia
Osteoporosis
Cardiovascular disorders, other
Musculoskeletal disorders, other
Cardiac valve disorders
Failure to thrive
Deficiency anemias
Respiratory disorders, other
Paralytic syndromes, other
Parkinsons disease
Spinal cord injury/disorders
Rheumatoid arthritis
Quadriplegia and paraplegia
Neurologic signs and symptoms
Autoimmune and connective tissue diseases

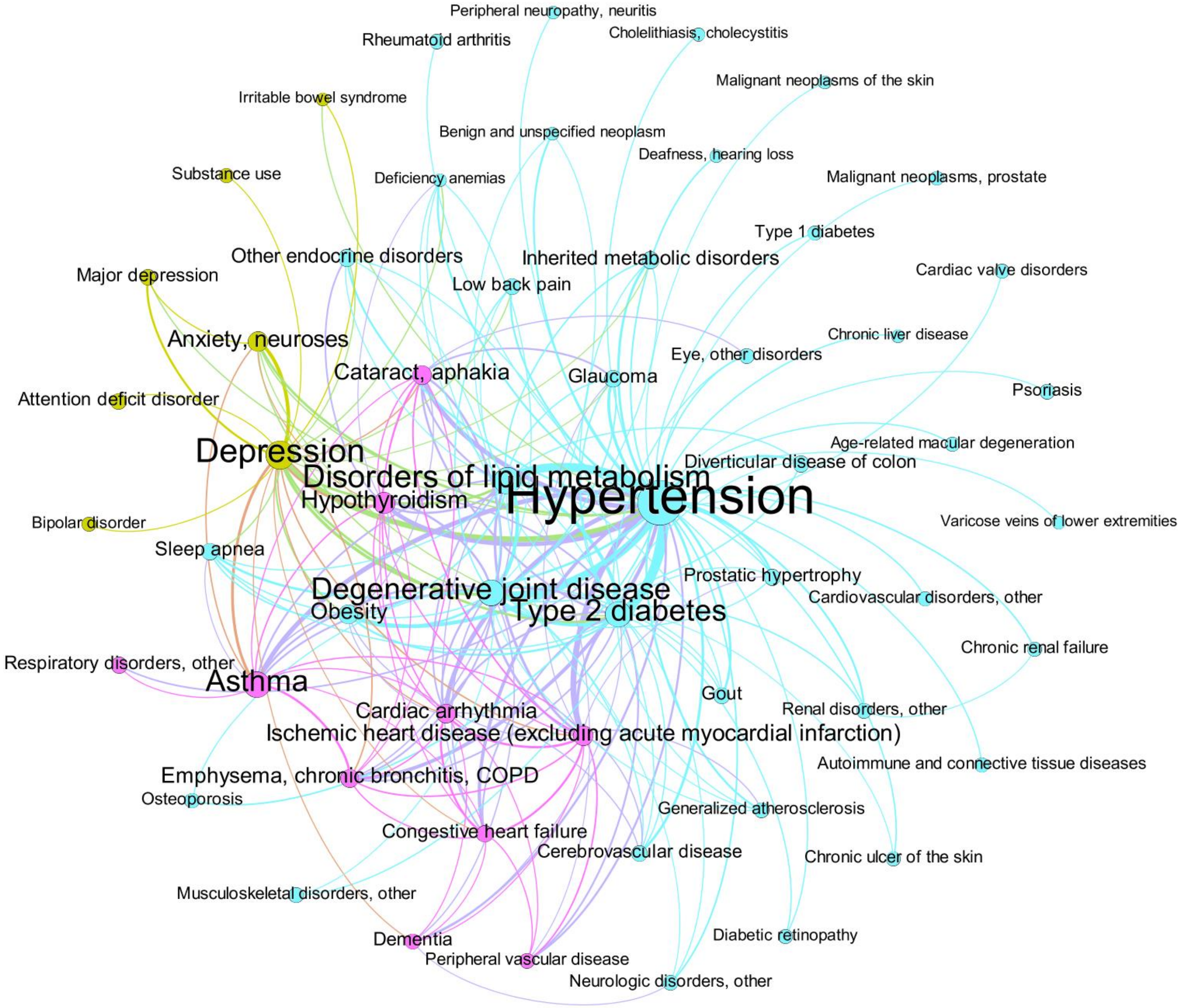
Note: Node diameter and font size are proportional to disease prevalence, edge weight (thickness) is proportional to effect size, and node and edge colour indicate community structure (i.e., disease clusters). COPD = chronic obstructive pulmonary disease, ESRD = end-stage renal disease.

Figure 10. Pairwise chronic disease network with co-occurrence relationships measured using Kulczynski and diseases ascertained using the Johns Hopkins ACG System, limited to the strongest 200 statistically significant associations.


Malignant neoplasms, prostate
Prostatic hypertrophy
Chronic liver disease
Prostatitis
Malignant neoplasms, stomach
Major depression
Genito-urinary disorders, other
Benign and unspecified neoplasm
Malignant neoplasms, esophagus
Malignant neoplasms, breast
Personality disorders
Malignant neoplasms, ovary
Acquired foot deformities
Deafness, hearing loss
Gastrointestinal signs and symptoms
Gastroesophageal reflux
Malignant neoplasms, cervix, uterus
Chronic pancreatitis
Malignant neoplasms, colorectal
Fungal infections
Renal calculi
Anxiety, neuroses
Incontinence
Glaucoma
Malignant neoplasms, lymphomas
Psych-physiologic and somatoform disorders
Inherited metabolic disorders
ENT disorders, other
Fluid/electrolyte disturbances
Malignant neoplasms, pancreas
Malignant neoplasms, bladder
Age-related macular degeneration
Hematologic disorders, other
Nephritis, nephrosis
Malignant neoplasms, kidney
Schizophrenia and affective psychosis
Gastrointestinal obstruction/perforation
Organic brain syndrome
Retinal disorders (excluding diabetic retinopathy)
Type 1 diabetes
Other endocrine disorders
Vesicoureteral reflux
Acute leukemia
Diverticular disease of colon
Eye, other disorders
Disorders of the immune system
Cataract, aphakia
Depression
Malignant neoplasms, liver and biliary tract
Malignant neoplasms, lung
Eating disorder
Cholelithiasis, cholecystitis
Diabetic retinopathy
Renal disorders, other
Blindness
Cardiac arrest, shock
Chronic ulcer of the skin
ESRD
Urinary symptoms
Substance use
Pulmonary embolism
Hypertension
Generalized atherosclerosis
Hypothyroidism
Impulse control
Disorders of mouth
Aortic aneurysm
Neurologic disorders, other
Aplastic anemia
Type 2 diabetes
Chronic renal failure
Malignant neoplasms of the skin
Disorders of the eyelid and lacrimal duct
Gout
Post traumatic stress disorder
Paralytic syndromes, other
Disorders of lipid metabolism
Cervical pain syndromes
Ophthalmic signs and symptoms
Acute myocardial infarction
Peripheral neuropathy, neuritis
Congestive heart failure
Peripheral vascular disease
Anorectal conditions
Dementia
Neurologic signs and symptoms
Psychological disorders of childhood
Nutritional deficiencies
Ischemic heart disease (excluding acute myocardial infarction)
Musculoskeletal signs and symptoms
Cerebrovascular disease
Low back pain
Emphysema, chronic bronchitis, COPD
Cardiac valve disorders
Parkinsons disease
Attention deficit disorder
Cardiomyopathy
Cardiovascular disorders, other
Bursitis, synovitis, tenosynovitis
Degenerative joint disease
Deficiency anemias
Cardiac arrhythmia
Musculoskeletal disorders, other
Delirium
Bipolar disorder
Acute respiratory failure
Cardiovascular signs and symptoms
Vertiginous syndromes
Varicose veins of lower extremities
Osteoporosis
Deep vein thrombosis
Gastrointestinal/hepatic disorders, other
Respiratory disorders, other
Asthma
Arthropathy
Obesity
Hemophilia, coagulation disorder
Spinal cord injury/disorders
Rheumatoid arthritis
Quadriplegia and paraplegia
Sleep apnea
Raynauds syndrome
Autoimmune and connective tissue diseases

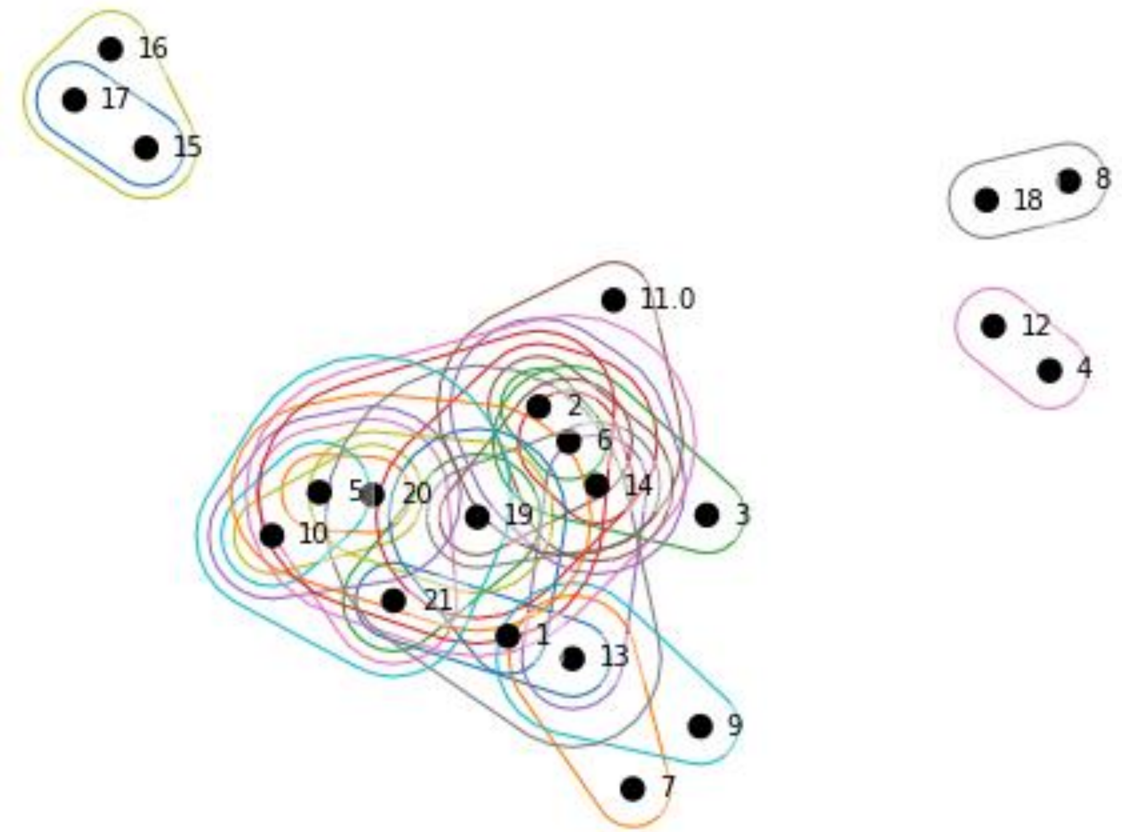
Note: Node diameter and font size are proportional to disease prevalence, edge weight (thickness) is proportional to effect size, and node and edge colour indicate community structure (i.e., disease clusters). COPD = chronic obstructive pulmonary disease; ENT = ear, nose, and throat; ESRD = end-stage renal disease.

Figure 11. Pairwise chronic disease network with co-occurrence relationships measured using joint prevalence and diseases ascertained using the Johns Hopkins ACG System, limited to the strongest 200 statistically significant associations.


Peripheral neuropathy, neuritis
Rheumatoid arthritis
Cholelithiasis, cholecystitis
Malignant neoplasms of the skin
Irritable bowel syndrome
Benign and unspecified neoplasm
Deafness, hearing loss
Substance use
Deficiency anemias
Malignant neoplasms, prostate
Type 1 diabetes
Other endocrine disorders
Inherited metabolic disorders
Major depression
Low back pain
Cardiac valve disorders
Anxiety, neuroses
Chronic liver disease
Eye, other disorders
Cataract, aphakia
Glaucoma
Psoriasis
Attention deficit disorder
Age-related macular degeneration
Depression
Diverticular disease of colon
Disorders of lipid metabolism
Hypertension
Hypothyroidism
Bipolar disorder
Varicose veins of lower extremities
Sleep apnea
Prostatic hypertrophy
Degenerative joint disease
Cardiovascular disorders, other
Obesity
Type 2 diabetes
Chronic renal failure
Respiratory disorders, other
Asthma
Gout
Renal disorders, other
Cardiac arrhythmia
Ischemic heart disease (excluding acute myocardial infarction)
Autoimmune and connective tissue diseases
Emphysema, chronic bronchitis, COPD
Osteoporosis
Generalized atherosclerosis
Congestive heart failure
Cerebrovascular disease
Chronic ulcer of the skin
Musculoskeletal disorders, other
Diabetic retinopathy
Dementia
Peripheral vascular disease
Neurologic disorders, other

Note: Node diameter and font size are proportional to disease prevalence, edge weight (thickness) is proportional to effect size, and node and edge colour indicate community structure (i.e., disease clusters). COPD = chronic obstructive pulmonary disease.

Figure 12. Chronic disease hypergraph constructed from the 30 strongest statistically significant pairwise and triad associations, with diseases ascertained using the Johns Hopkins ACG System and co-occurrence relationships measured using phi.

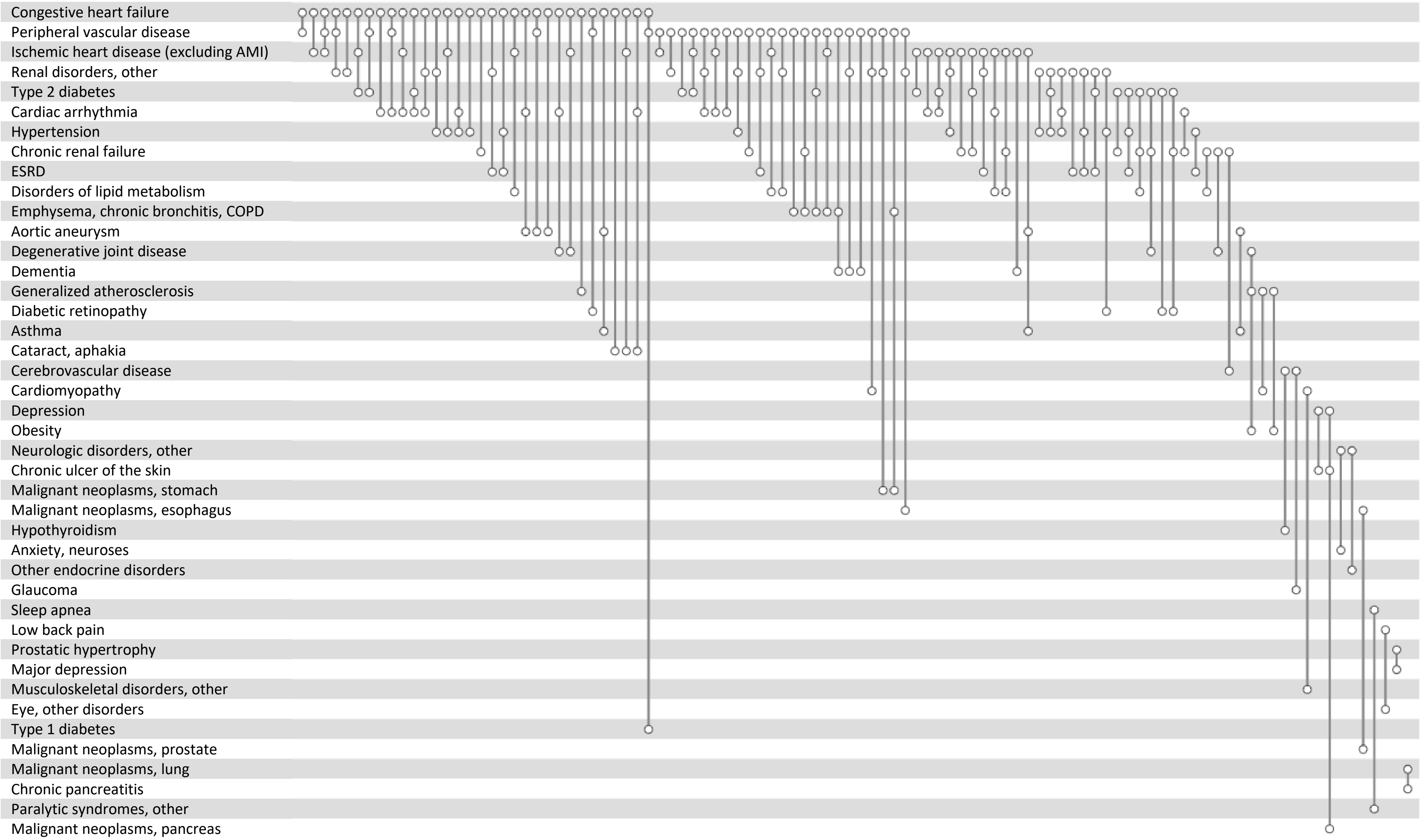


Legend	
1	Aortic aneurysm
2	Cardiac arrhythmia
3	Cardiomyopathy
4	Cataract, aphakia
5	Chronic renal failure
6	Congestive heart failure
7	Degenerative joint disease
8	Dementia
9	Disorders of lipid metabolism
10	ESRD
11	Emphysema, chronic bronchitis, COPD
12	Glaucoma
13	Hypertension
14	Ischemic heart disease (excluding AMI)
15	Malignant neoplasms, esophagus
16	Malignant neoplasms, lung
17	Malignant neoplasms, stomach
18	Neurologic disorders, other
19	Peripheral vascular disease
20	Renal disorders, other
21	Type 2 diabetes

Note: Coloured bounding containers (hyperedges) indicate co-occurrence relationships amongst chronic disease nodes. AMI = acute myocardial infarction, COPD = chronic obstructive pulmonary disease, ESRD = end-stage renal disease.

Figure 13. Parallel Aggregated Ordered Hypergraph (PAOH) visualization of a chronic disease hypergraph constructed from the 100 strongest statistically significant pairwise and triad associations, with diseases ascertained using the Johns Hopkins ACG System and co-occurrence relationships measured using phi.

Note: AMI = acute myocardial infarction, COPD = chronic obstructive pulmonary disease, ESRD = end-stage renal disease.



4.5 Co-occurrence Relationships Characterized by Disease Prevalence

Different co-occurrence measures estimate higher association strengths for different types of relationships, in terms of the prevalence difference between disease pairs. These preferences by association measures result in certain pairwise chronic disease relationships being emphasized more than other disease combinations, when limiting networks to the strongest associations. Differences based on disease prevalence were more pronounced when using a smaller number of the strongest associations (Figure 14, Figure 16, Table 8) and decreased when including a larger number of all measured associations (Figure 15, Figure 16, Table 9). The overall patterns remained consistent while percentages varied for the sensitivity analysis, in which prevalence was classified as low (<0.5%), moderate (0.5 to <5%), and high (≥5%) (Appendix F, Appendix G, Appendix H, Appendix I).

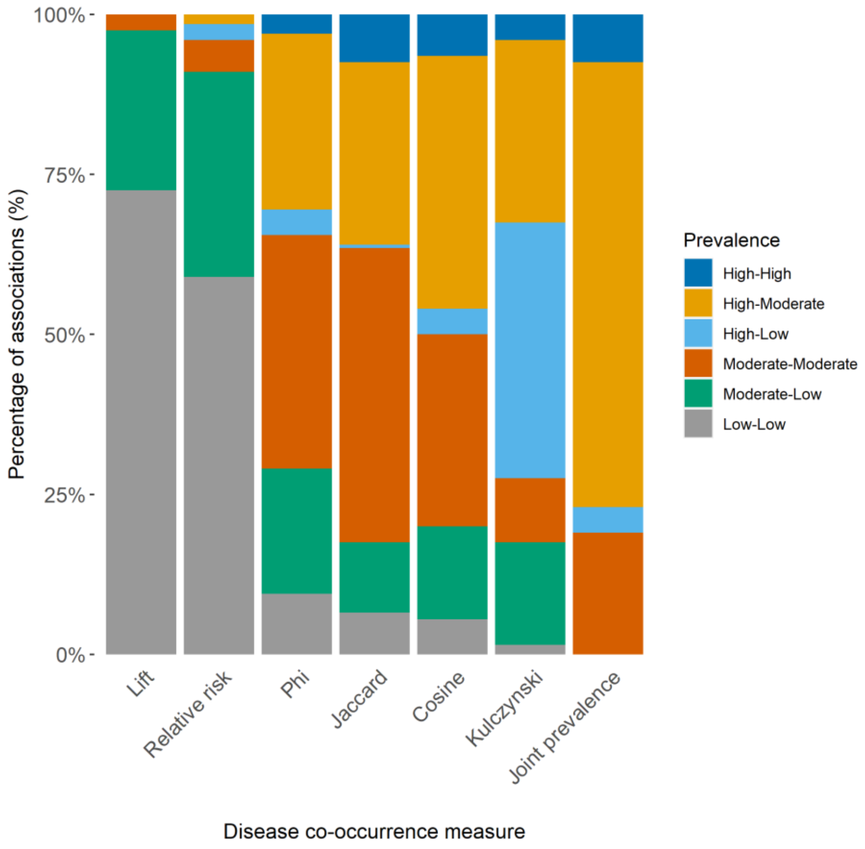
Networks based on lift and relative risk accentuated co-occurrence relationships between pairs of low prevalent (<1%) conditions, at 72.5% and 59.0% respectively (Figure 14, Table 8). The percentage of edges highlighting co-occurrences between two low prevalent conditions in the other five networks ranged from 0% (joint prevalence) to 9.5% (phi). Lift and relative risk also highlighted a higher proportion of relationships between moderately prevalent (1 to <7%) and low prevalent conditions, compared with the other co-occurrence measures.

Relationships between two moderately prevalent conditions were emphasized more by phi, Jaccard, and cosine based networks: 36.5%, 46.0%, 30.0%, respectively. Phi, Jaccard, and cosine also emphasized relationships between highly and moderately prevalent diseases: 27.5%, 28.5%, 39.5%. The majority of the edges in the Kulczynski-based network represented relationships between conditions of high and low prevalence (40.0%), and between highly prevalent and moderately prevalent conditions (28.5%). Relationships between conditions of high and low prevalence only constituted up to 4.0% of all edges in the other six networks.

Measuring co-occurrence using joint prevalence resulted in the highest percentage of edges connecting highly prevalent and moderately prevalent disease nodes (69.5%). Joint prevalence and Jaccard, resulted in the most connections between two highly prevalent conditions (7.5%). Correspondingly, the joint prevalence network had the highest median joint prevalence (0.7%, Q1-Q3: 0.6%-1.2%) (Table 10). Lift and relative risk based networks did not contain any edges between two highly prevalent disease nodes, while associations between pairs of highly prevalent conditions accounted for 3.0% to 6.5% of the edges in networks built using phi, cosine, and Kulczynski.

The median difference in prevalence between pairs of co-occurring conditions was lowest for lift (0.3%, Q1-Q3: 0.1-0.8%) and relative risk (0.4%, Q1-Q3: 0.1-1.3%); and highest for Kulczynski (17.9%, Q1-Q3: 3.6-22.0%) (Table 10). There was less variation in the distribution of prevalence differences among the seven co-occurrence measures when 50% of all statistically significant associations were included (Table 11).

Figure 14. Percentage of the strongest 200 statistically significant pairwise chronic disease co-occurrence relationships characterized by prevalence, among select co-occurrence measures.



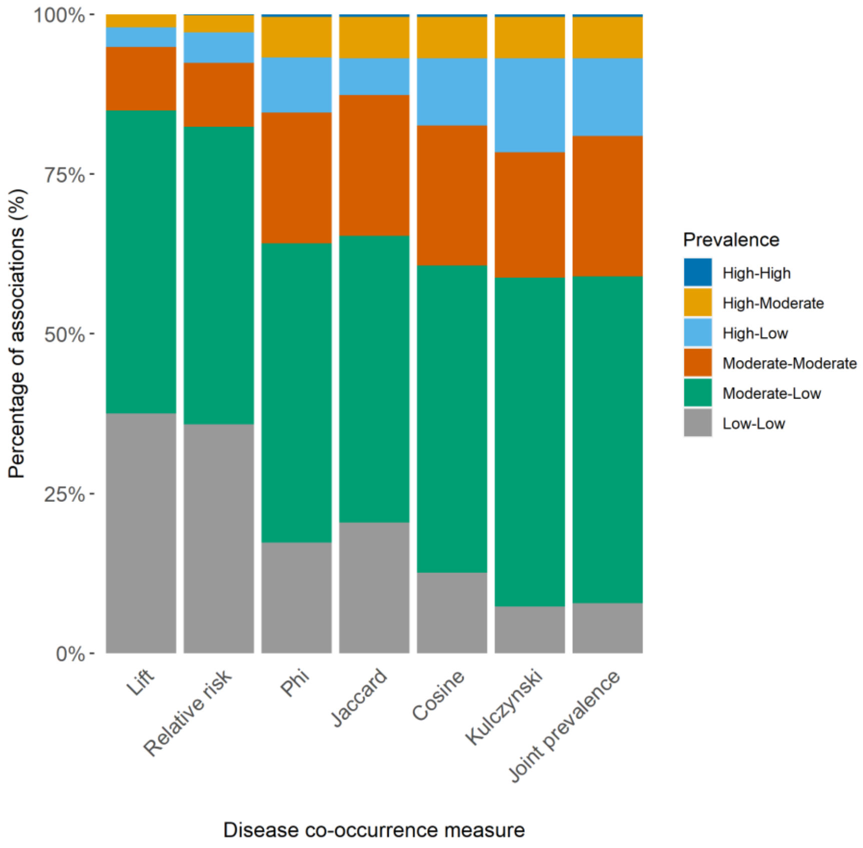
Note: Chronic diseases were ascertained using the Johns Hopkins ACG System; and prevalence was categorized as low (<1%), moderate (1 to <7%), and high (≥7%).

Table 8. Number (%) of the strongest 200 statistically significant pairwise chronic disease co-occurrence relationships characterized by prevalence, among select co-occurrence measures.

Prevalence	Lift	Relative risk	Phi	Jaccard	Cosine	Kulczynski	Joint prevalence
High-High	0 (0.0)	0 (0.0)	6 (3.0)	15 (7.5)	13 (6.5)	8 (4.0)	15 (7.5)
High-Moderate	0 (0.0)	3 (1.5)	55 (27.5)	57 (28.5)	79 (39.5)	57 (28.5)	139 (69.5)
High-Low	0 (0.0)	5 (2.5)	8 (4.0)	1 (0.5)	8 (4.0)	80 (40.0)	8 (4.0)
Moderate-Moderate	5 (2.5)	10 (5.0)	73 (36.5)	92 (46.0)	60 (30.0)	20 (10.0)	38 (19.0)
Moderate-Low	50 (25.0)	64 (32.0)	39 (19.5)	22 (11.0)	29 (14.5)	32 (16.0)	0 (0.0)
Low-Low	145 (72.5)	118 (59.0)	19 (9.5)	13 (6.5)	11 (5.5)	3 (1.5)	0 (0.0)

Note: Chronic diseases were ascertained using the Johns Hopkins ACG System; prevalence was categorized as low (<1%), moderate (1 to <7%), and high (≥7%).

Figure 15. Percentage of the strongest 50 percent (*n*=3,922) of statistically significant pairwise chronic disease co-occurrence relationships characterized by prevalence, among select co-occurrence measures.



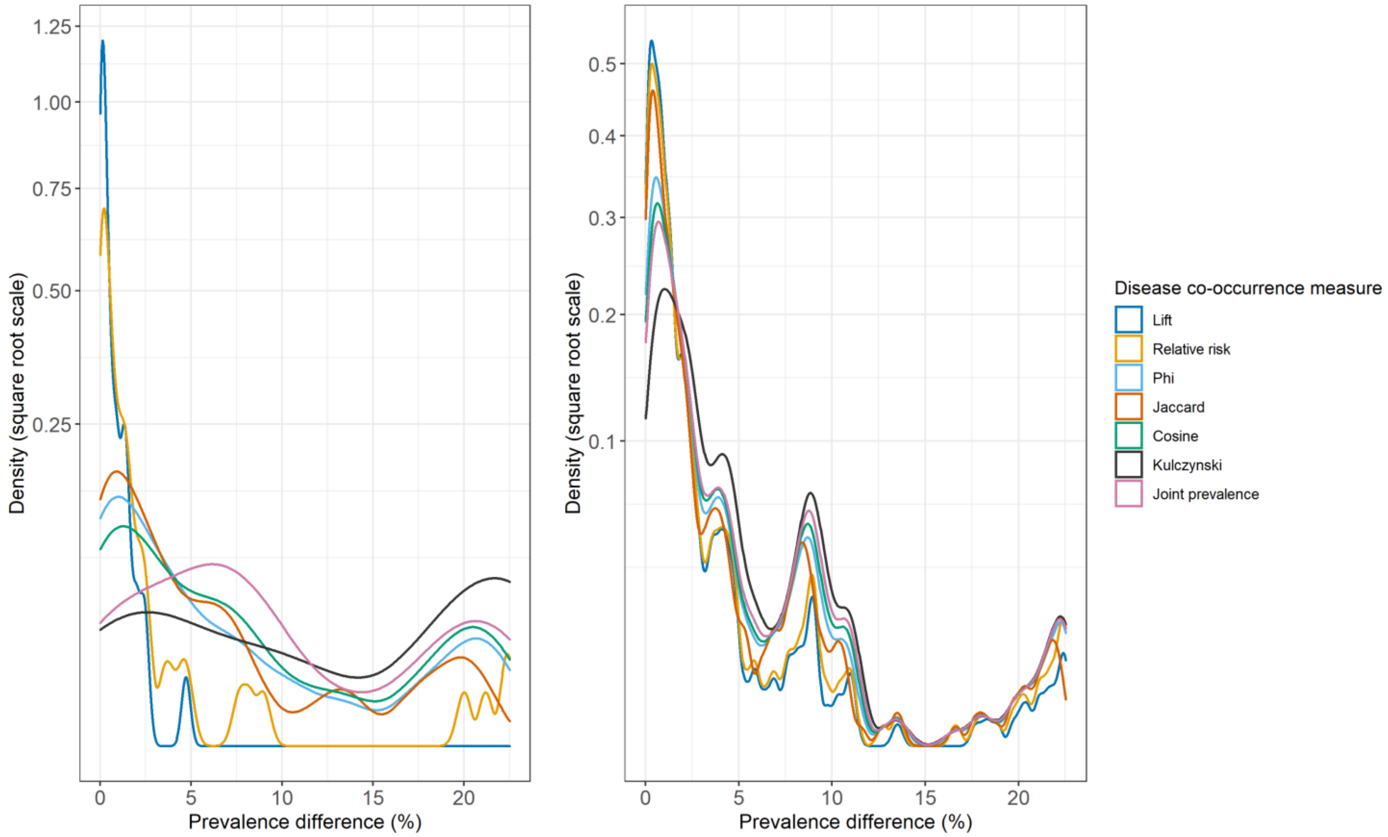
Note: Chronic diseases were ascertained using the Johns Hopkins ACG System; and prevalence was categorized as low (<1%), moderate (1 to <7%), and high (≥7%).

Table 9. Number (%) of the strongest 50 percent (*n*=3,922) of statistically significant pairwise chronic disease co-occurrence relationships characterized by prevalence, among select co-occurrence measures.

Prevalence	Lift	Relative risk	Phi	Jaccard	Cosine	Kulczynski	Joint prevalence
High-High	1 (0.0)	6 (0.2)	15 (0.4)	15 (0.4)	15 (0.4)	15 (0.4)	15 (0.4)
High-Moderate	78 (2.0)	105 (2.7)	250 (6.4)	255 (6.5)	255 (6.5)	255 (6.5)	255 (6.5)
High-Low	122 (3.1)	188 (4.8)	339 (8.6)	225 (5.7)	414 (10.6)	577 (14.7)	476 (12.1)
Moderate-Moderate	389 (9.9)	392 (10.0)	801 (20.4)	864 (22.0)	857 (21.9)	769 (19.6)	864 (22.0)
Moderate-Low	1,861 (47.5)	1,827 (46.6)	1,837 (46.8)	1,761 (44.9)	1,887 (48.1)	2,020 (51.5)	2,004 (51.1)
Low-Low	1,471 (37.5)	1,404 (35.8)	680 (17.3)	802 (20.4)	494 (12.6)	286 (7.3)	308 (7.9)

Note: Chronic diseases were ascertained using the Johns Hopkins ACG System; prevalence was categorized as low (<1%), moderate (1 to <7%), and high (≥7%).

Figure 16. Prevalence difference between pairs of co-occurring chronic conditions in pairwise networks limited to the strongest 200 statistically significant associations (left) and limited to the strongest 50 percent (*n*=3,922) of all statistically significant associations (right), among select co-occurrence measures.



Note: Chronic diseases were ascertained using the Johns Hopkins ACG System.

Table 10. Summary of effect size, joint prevalence, and prevalence difference distributions among pairwise chronic disease co-occurrence networks constructed from the strongest 200 statistically significant associations, among select co-occurrence measures.

	Lift	Relative risk	Phi	Jaccard	Cosine	Kulczynski	Joint prevalence
Effect size							
Median (Q1-Q3)	23.4 (17.9-33.4)	29.5 (22.6-46.1)	0.2 (0.1-0.2)	0.1 (0.1-0.1)	0.2 (0.2-0.2)	0.3 (0.2-0.4)	0.7 (0.6-1.2)
Range	15.9-405.0	19.3-8,627.8	0.1-0.4	0.1-0.3	0.1-0.5	0.2-0.5	0.4-6.5
Joint prevalence							
Median (Q1-Q3)	0.0 (0.0-0.0)	0.0 (0.0-0.1)	0.4 (0.2-0.9)	0.5 (0.3-1.1)	0.6 (0.3-1.2)	0.4 (0.1-0.9)	0.7 (0.6-1.2)
Range	0.0-0.6	0.0-2.2	0.0-6.5	0.0-6.5	0.0-6.5	0.0-6.5	0.4-6.5
Prevalence difference							
Median (Q1-Q3)	0.3 (0.1-0.8)	0.4 (0.1-1.3)	2.2 (0.6-6.3)	2.0 (0.5-5.0)	3.2 (1.1-7.7)	17.9 (3.6-22.0)	6.8 (3.4-12.9)
Range	0.0-4.7	0.0-22.5	0.0-21.7	0.0-20.6	0.0-21.8	0.0-22.5	0.0-21.8

Note: Chronic diseases were ascertained using the Johns Hopkins ACG System.

Table 11. Summary of effect size, joint prevalence, and prevalence difference distributions among pairwise chronic disease co-occurrence networks constructed from the strongest 50 percent (*n*=3,922) of statistically significant associations, among select co-occurrence measures.

	Lift	Relative risk	Phi	Jaccard	Cosine	Kulczynski	Joint prevalence
Effect size							
Median (Q1-Q3)	4.1 (3.2-6.2)	4.5 (3.5-7.0)	0.0 (0.0-0.0)	0.0 (0.0-0.0)	0.0 (0.0-0.1)	0.1 (0.0-0.1)	0.0 (0.0-0.1)
Range	2.7-405.0	2.9-8,627.8	0.0-0.4	0.0-0.3	0.0-0.5	0.0-0.5	0.0-6.5
Joint prevalence							
Median (Q1-Q3)	0.0 (0.0-0.0)	0.0 (0.0-0.0)	0.0 (0.0-0.1)	0.0 (0.0-0.1)	0.0 (0.0-0.1)	0.0 (0.0-0.1)	0.0 (0.0-0.1)
Range	0.0-6.5	0.0-6.5	0.0-6.5	0.0-6.5	0.0-6.5	0.0-6.5	0.0-6.5
Prevalence difference							
Median (Q1-Q3)	0.9 (0.4-1.9)	1.0 (0.4-2.1)	1.4 (0.6-3.4)	1.1 (0.4-2.4)	1.5 (0.6-3.8)	2.3 (1.1-4.7)	1.7 (0.7-4.1)
Range	0.0-22.5	0.0-22.5	0.0-22.5	0.0-22.3	0.0-22.5	0.0-22.5	0.0-22.5

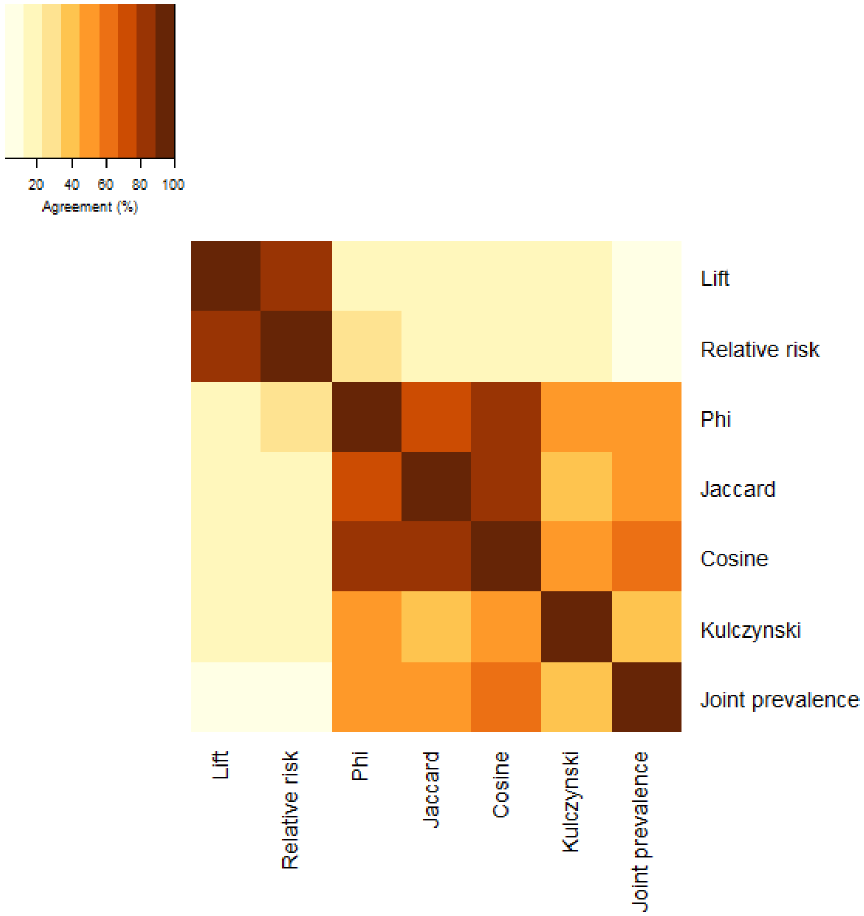
Note: Chronic diseases were ascertained using the Johns Hopkins ACG System.

4.6 Network Edge Similarity

Disease networks constructed using different co-occurrence measures were dissimilar in terms of the edges included in the top 200 associations (Figure 17, Table 12). Edge agreement ranged from 1.5% for lift and joint prevalence to 86.5% for lift and relative risk. Phi- and Jaccard-based networks had moderate agreement with the cosine-based network (83.0% and 79.5%). Phi and Jaccard had moderate agreement (78.0%), while the remaining network pairs had lower agreement; it ranged from 5.0% to 63.5%. Median agreement (37.0%, Q1-Q3: 20.0%-53.5%) among the network pairs was much lower when limited to the strongest 200 associations, than when the top 50 percent of all statistically significant associations were used to construct the networks (68.5%, Q1-Q3: 58.7%-83.9%) (Figure 18, Table 13).

When comparing the strongest 200 associations between the pairwise networks and their respective hypergraphs, by extracting the binary relationships from the higher-order hyperedges, the percentage of pairwise network associations also represented within the respective hypergraph ranged from 28.5% (lift) to 74.5% (Jaccard) and the median agreement was 57.0% (Q1-Q3: 41.0%-65.0%).

Figure 17. Percent (%) of the strongest 200 statistically significant pairwise chronic disease co-occurrence relationships in common between networks constructed using different co-occurrence measures.



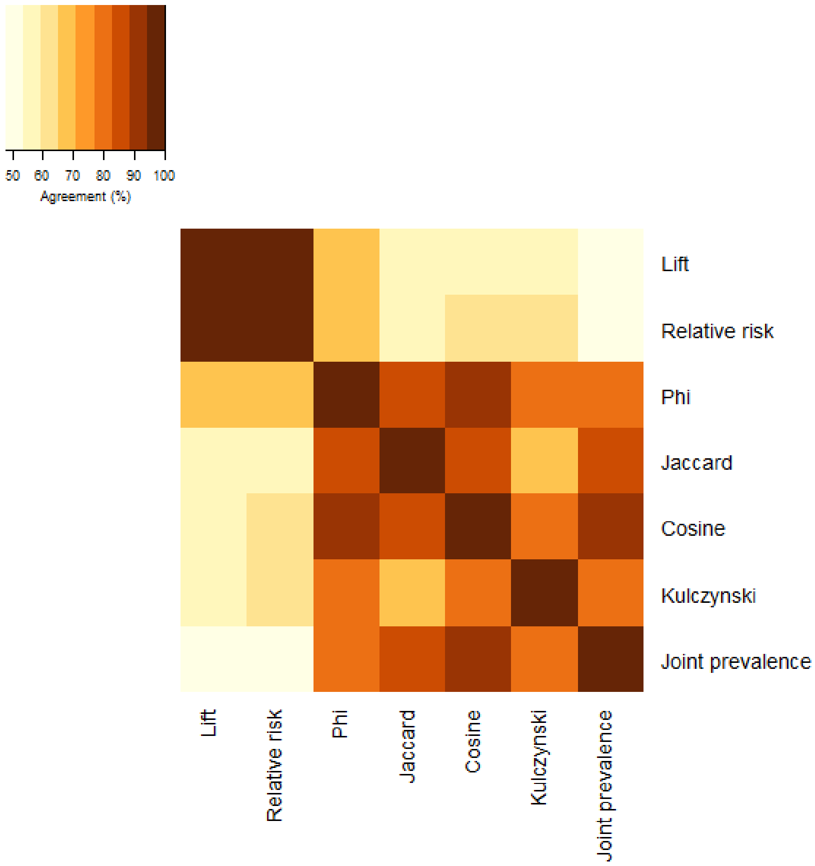
Note: Chronic diseases were ascertained using the Johns Hopkins ACG System.

Table 12. Percent (%) of the strongest 200 statistically significant pairwise chronic disease co-occurrence relationships in common between networks constructed using different co-occurrence measures.

	Lift	**Relative risk**	**Phi**	**Jaccard**	**Cosine**	**Kulczynski**	**Joint prevalence**
Lift	100.0						
Relative risk	86.5	100.0					
Phi	22.5	29.5	100.0				
Jaccard	15.0	20.0	78.0	100.0			
Cosine	16.5	23.0	83.0	79.5	100.0		
Kulczynski	14.0	22.5	50.0	37.0	52.0	100.0	
Joint prevalence	1.5	5.0	46.5	53.5	63.5	44.5	100.0

Note: Chronic diseases were ascertained using the Johns Hopkins ACG System.

Figure 18. Percent (%) of the strongest 50 percent (*n*=3,922) of all statistically significant pairwise chronic disease co-occurrence relationships in common between networks constructed using different co-occurrence measures.



Note: Chronic diseases were ascertained using the Johns Hopkins ACG System.

Table 13. Percent (%) of the strongest 50 percent (*n*=3,922) of all statistically significant pairwise chronic disease co-occurrence relationships in common between networks constructed using different co-occurrence measures.

	Lift	**Relative risk**	**Phi**	**Jaccard**	**Cosine**	**Kulczynski**	**Joint prevalence**
Lift	100.0						
Relative risk	96.8	100.0					
Phi	66.6	68.5	100.0				
Jaccard	55.5	56.5	83.9	100.0			
Cosine	57.6	59.5	91.0	87.1	100.0		
Kulczynski	58.7	61.5	80.3	68.0	80.6	100.0	
Joint prevalence	47.7	49.6	81.1	85.7	90.0	77.2	100.0

Note: Chronic diseases were ascertained using the Johns Hopkins ACG System.

4.7 Community Structure

Community structure differed considerably amongst networks constructed using different co-occurrence measures. The number of communities (i.e., clusters) detected had the largest range (3 to 17) between networks limited to 200 of all statistically significant associations (Table 14). However, networks containing 50 percent of all EDC-based associations (2 to 6), all EDC-based associations (3 to 7), and based on all Elixhauser-based associations (2 to 5) also had considerable dissimilarity in the number of communities detected. Modularity, a measure of how well a network separates into communities, also widely varied between networks constructed using different co-occurrence measures. Variation in modularity between the networks decreased, as more associations were included. When all EDC associations were included, modularity ranged from 0.07 (joint prevalence) to 0.36 (relative risk) for the pairwise networks, but no community structure was identified in any of the seven hypergraphs that incorporated triad associations (modularity=0).

Table 14. Community structure properties for networks constructed with the strongest 200 and strongest 50 percent (*n*=3,922) of all statistically significant pairwise chronic disease co-occurrence relationships measured using the Johns Hopkins ACG System, and all statistically significant pairwise co-occurrences measured using the Elixhauser index.

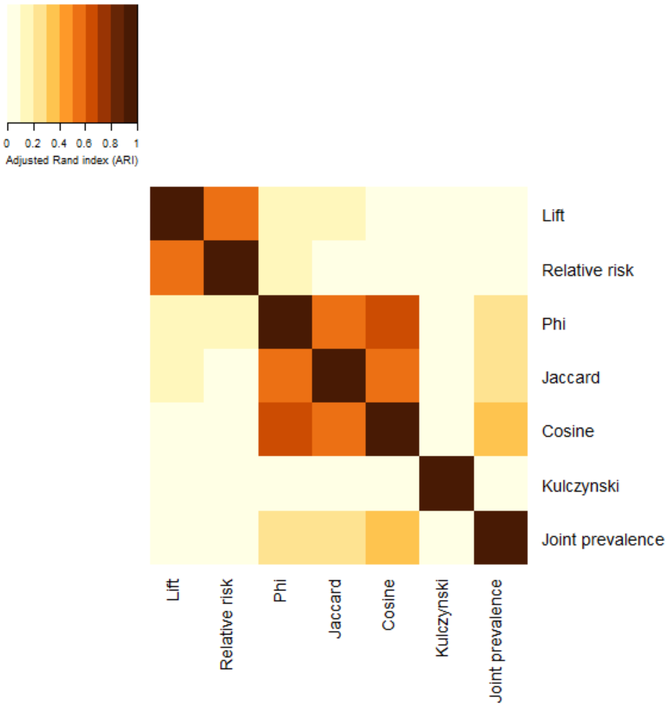
Association measure	**Top 200 associations**		**Top 50 percent of associations**		**All associations (Elixhauser index)**	
	Modularity	N communities	Modularity	N communities	Modularity	N communities
Lift	0.72	13	0.30	6	0.09	5
Relative risk	0.60	13	0.43	5	0.21	4
Phi	0.43	17	0.19	4	0.11	3
Jaccard	0.37	14	0.16	5	0.11	4
Cosine	0.37	11	0.15	4	0.07	3
Kulczynski	0.37	8	0.14	5	0.08	3
Joint prevalence	0.08	3	0.07	2	0.03	2

Community structure similarity, as measured using the adjusted Rand index, was strongest between phi and cosine in networks limited to the top 200 associations (ARI=0.68) (Figure 19, Table 15). The strongest similarity among networks limited to the top 50 percent of associations, was between relative risk and lift (ARI=0.49) and between phi and cosine

(ARI=0.48) (Figure 20, Table 16). Phi and Kulczynski had perfect agreement (ARI=1) in networks constructed using all associations based on the Elixhauser index (Figure 21, Table 17).

Overall, co-occurrence measurement differences resulted in poor similarity: the median ARI was 0.08 (Q1-Q3: 0.06-0.24) for networks including the top 200 associations, and the median was 0.26 (Q1-Q3: 0.24-0.32) for networks consisting of the top 50 percent of associations. When all statistically significant associations (disease ascertainment using the Elixhauser index algorithms) were included, the median ARI was 0.38 (Q1-Q3: 0.28-0.67). Similarities and differences in community structure between relative risk and Jaccard-based chronic disease networks are shown in Figure 22.

Figure 19. Community structure similarity, measured using the adjusted Rand index (ARI), between chronic disease networks constructed using different co-occurrence measures and limited to the strongest 200 statistically significant pairwise relationships.



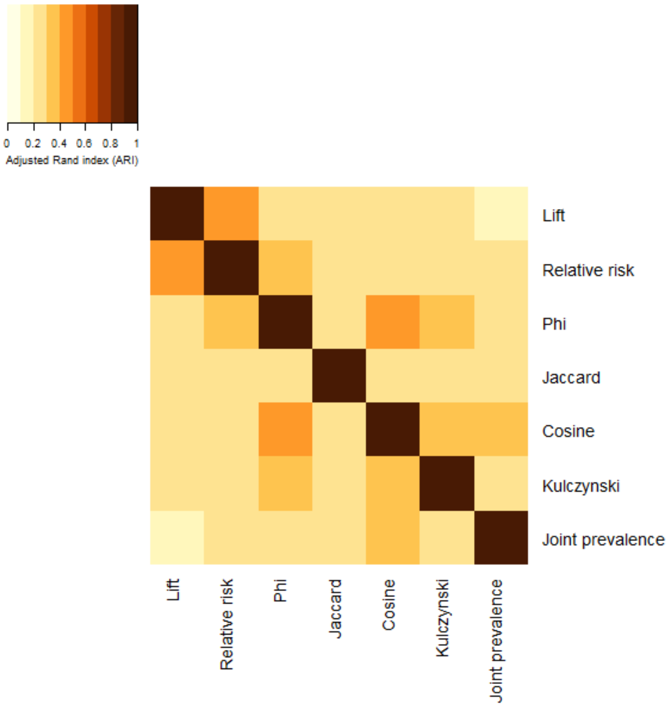
Note: Chronic diseases were ascertained using the Johns Hopkins ACG System.

Table 15. Community structure similarity, measured using the adjusted Rand index (ARI), between chronic disease networks constructed using different co-occurrence measures and limited to the strongest 200 statistically significant pairwise relationships.

	Lift	**Relative risk**	**Phi**	**Jaccard**	**Cosine**	**Kulczynski**	**Joint prevalence**
Lift	1.00						
Relative risk	0.58	1.00					
Phi	0.12	0.10	1.00				
Jaccard	0.18	0.06	0.52	1.00			
Cosine	0.08	0.08	0.68	0.54	1.00		
Kulczynski	0.08	0.06	0.07	0.05	0.07	1.00	
Joint prevalence	-0.01	0.00	0.23	0.24	0.40	0.05	1.00

Note: Chronic diseases were ascertained using the Johns Hopkins ACG System.

Figure 20. Community structure similarity, measured using the adjusted Rand index (ARI), between chronic disease networks constructed using different co-occurrence measures and limited to the strongest 50 percent (*n*=3,922) of all statistically significant pairwise relationships.



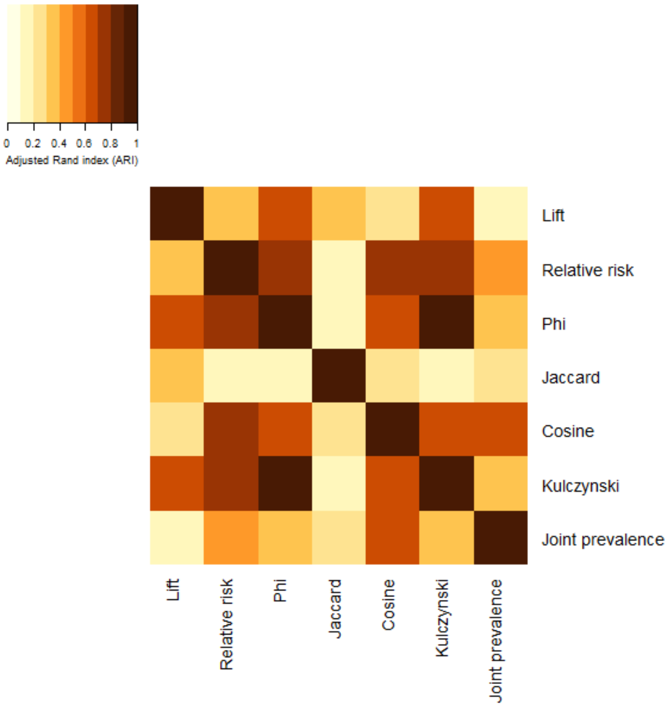
Note: Chronic diseases were ascertained using the Johns Hopkins ACG System.

Table 16. Community structure similarity, measured using the adjusted Rand index (ARI), between chronic disease networks constructed using different co-occurrence measures and limited to the strongest 50 percent (*n*=3,922) of all statistically significant pairwise relationships.

	Lift	Relative risk	Phi	Jaccard	Cosine	Kulczynski	Joint prevalence
Lift	1.00						
Relative risk	0.49	1.00					
Phi	0.28	0.33	1.00				
Jaccard	0.26	0.21	0.28	1.00			
Cosine	0.25	0.27	0.48	0.26	1.00		
Kulczynski	0.29	0.21	0.32	0.20	0.40	1.00	
Joint prevalence	0.20	0.24	0.25	0.24	0.34	0.21	1.00

Note: Chronic diseases were ascertained using the Johns Hopkins ACG System.

Figure 21. Community structure similarity, measured using the adjusted Rand index (ARI), between chronic disease networks constructed using different co-occurrence measures, including all statistically significant pairwise relationships.



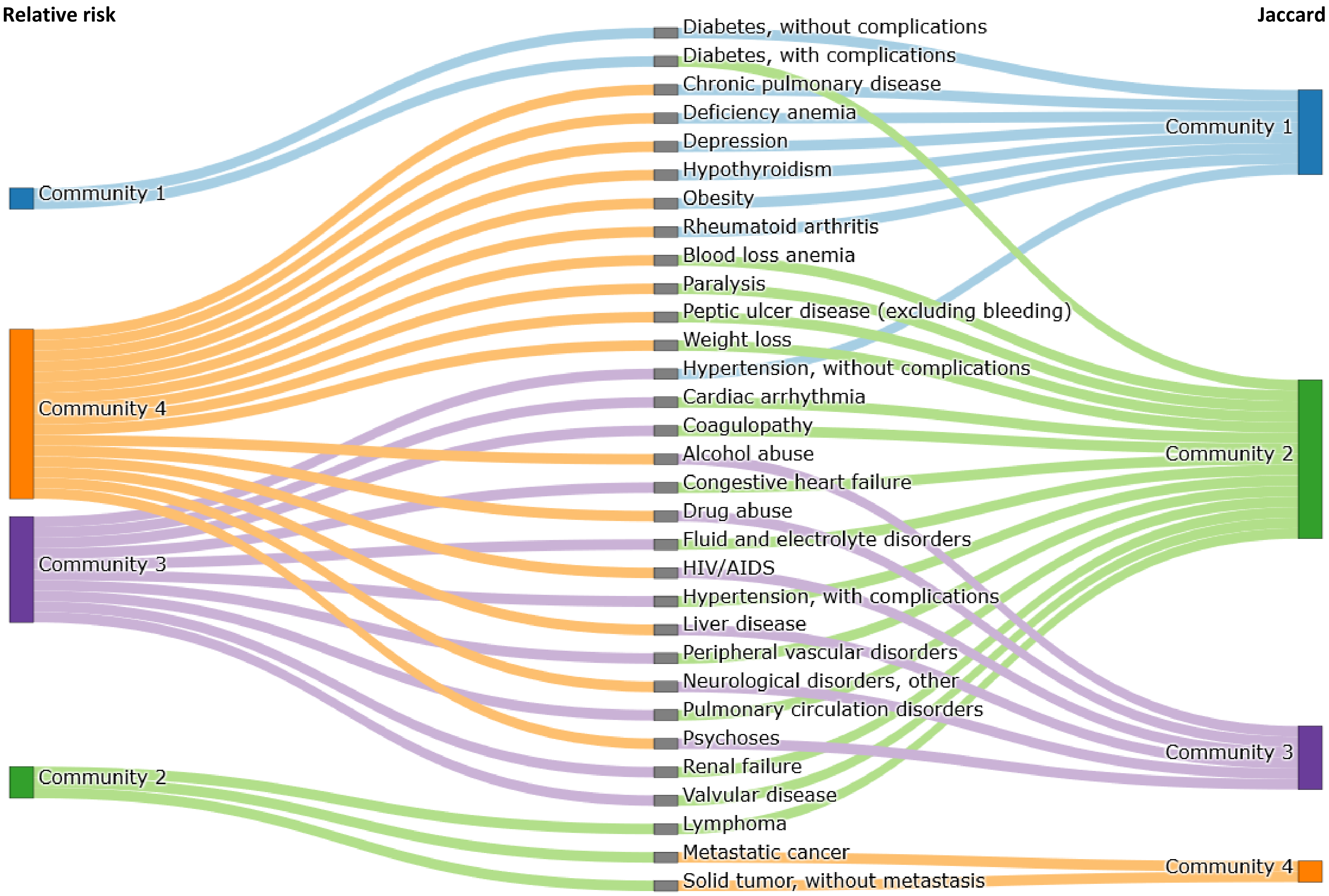
Note: Chronic diseases were ascertained using the Elixhauser comorbidity index.

Table 17. Community structure similarity, measured using the adjusted Rand index (ARI), between chronic disease networks constructed using different co-occurrence measures, including all statistically significant pairwise relationships.

	Lift	Relative risk	Phi	Jaccard	Cosine	Kulczynski	Joint prevalence
Lift	1.00						
Relative risk	0.38	1.00					
Phi	0.63	0.79	1.00				
Jaccard	0.32	0.18	0.19	1.00			
Cosine	0.30	0.77	0.67	0.28	1.00		
Kulczynski	0.63	0.79	1.00	0.19	0.67	1.00	
Joint prevalence	0.20	0.47	0.37	0.22	0.67	0.37	1.00

Note: Chronic diseases were ascertained using the Elixhauser comorbidity index.

Figure 22. Comparison of community structure for chronic disease networks based on the Elixhauser comorbidity index with disease co-occurrence measured using relative risk (left) and Jaccard (right).



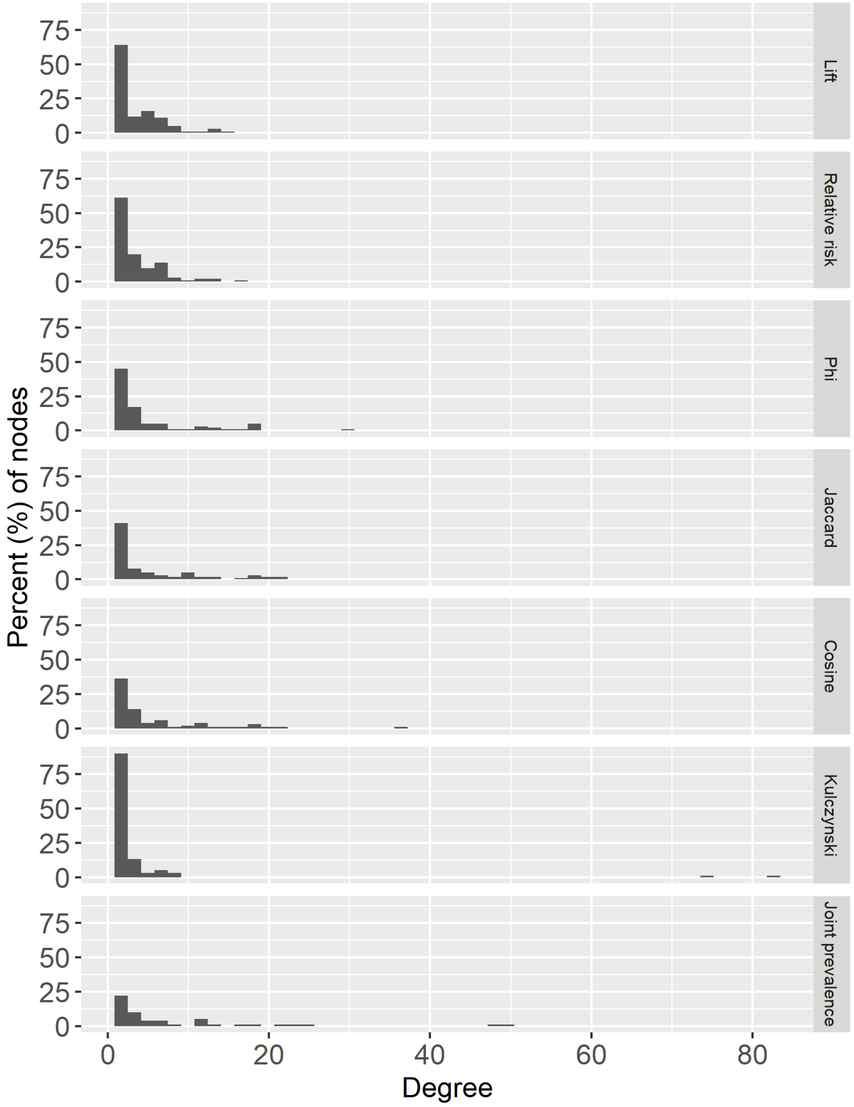
4.8 Nodes of Importance

Since degree centrality is a non-weighted measure, networks that included all statistically significant edges, without limiting inclusion by effect size, had identical degree distributions. When network complexity was reduced by excluding edges by effect size to create a visually interpretable network diagram, degree distribution varied considerably amongst pairwise networks constructed using different co-occurrence measures (Figure 23, Figure 24).

The selection of the top 20 disease categories with the highest degree centrality varied amongst networks constructed using different co-occurrence measures. Agreement between the networks limited to the top 200 co-occurrence relationships varied, with a median of 55.0% (Q1-Q3: 25.0%-75.0%, Figure 25) and a median of 55.0% (Q1-Q3: 30.0%-75.0%, Figure 26) when limited to the strongest 50 percent of associations. When limited to the top 200 co-occurrences, agreement ranged from 5% between lift and joint prevalence to 95% between Jaccard and cosine. Agreement between two of the most commonly used measures among disease network studies, relative risk and phi, agreed on only 30% of the top 20 central nodes. When 50 percent of all statistically significant associations were included, agreement was strongest between Kulczynski and joint prevalence (95% agreement), and weakest between lift and Kulczynski (20%) and between lift and joint prevalence (20%). Table 18 compares the top 20 disease nodes with the highest degree centrality (i.e., most commonly co-occurring with other conditions) among networks limited to the top 200 co-occurrences measured using phi, relative risk, and joint prevalence.

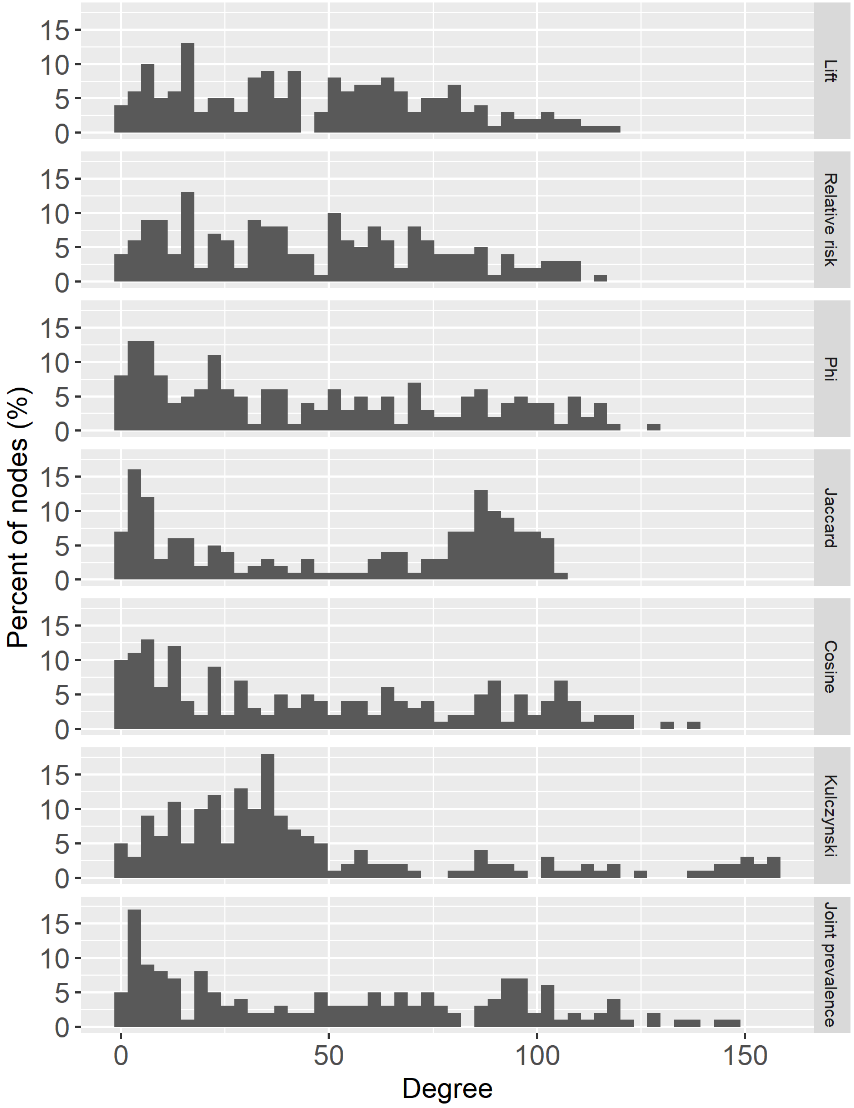
When including all statistically significant associations, the five chronic disease categories with the highest degree centrality in the pairwise network were “other endocrine disorders,” depression, major depression, sleep apnea, and asthma (Table 19). Meanwhile, the five most central nodes in the hypergraph built using both dyad and triad associations were hypertension, degenerative joint disease, depression, type 2 diabetes, and ischemic heart disease (excluding acute myocardial infarction). The pairwise network and the hypergraph had poor agreement (20%) when considering the top 10 most central nodes and moderate agreement (65%) when comparing the top 20 most central nodes.

Figure 23. Node degree distribution for chronic disease networks constructed using select co-occurrence measures and limited to the strongest 200 statistically significant pairwise associations.



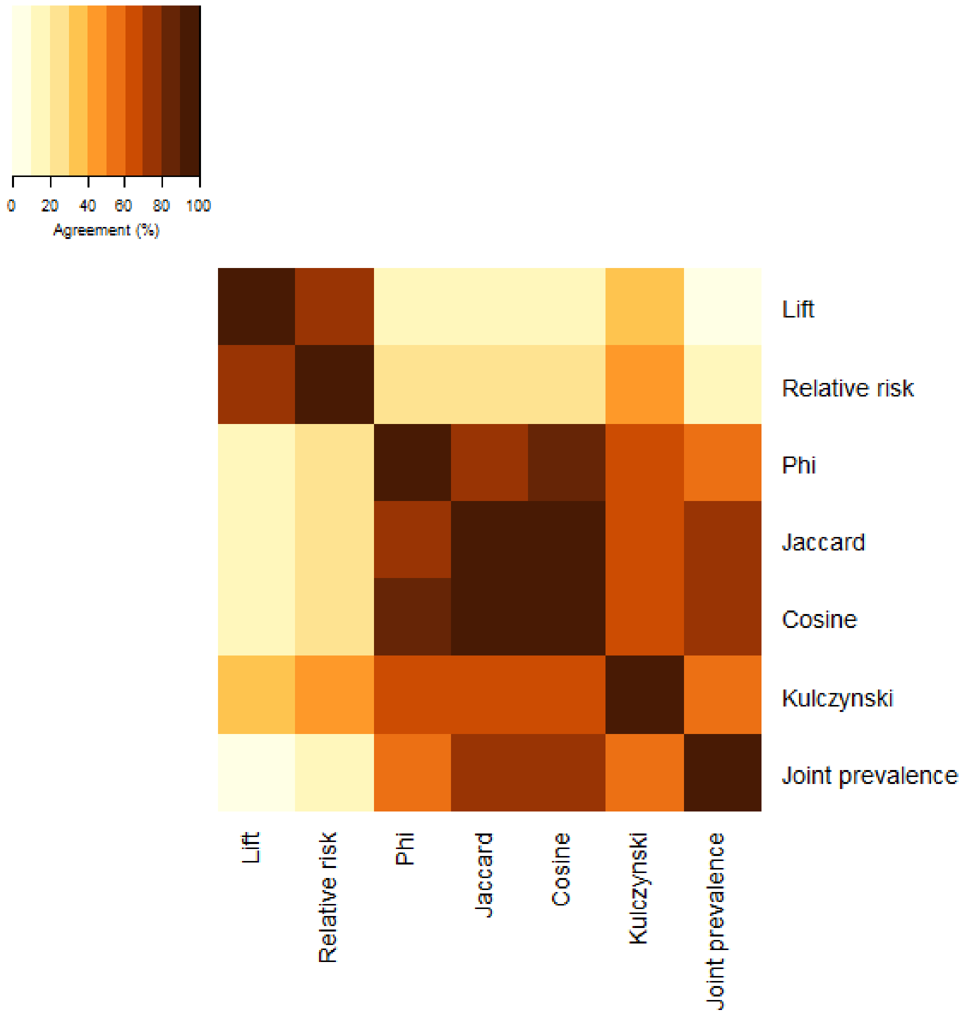
Note: Chronic diseases were ascertained using the Johns Hopkins ACG System.

Figure 24. Node degree distribution for chronic disease networks constructed using select co-occurrence measures and limited to the strongest 50 percent (*n*=3,922) of all statistically significant pairwise associations.



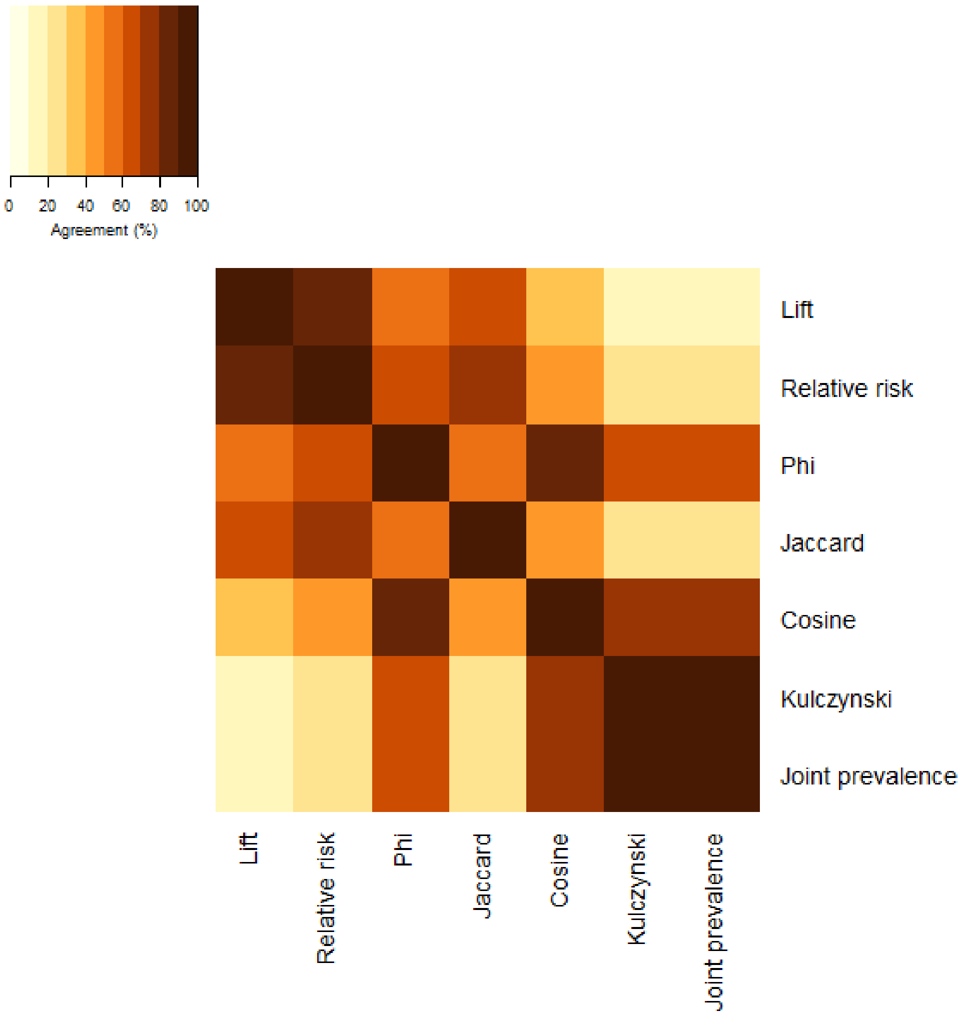
Note: Chronic diseases were ascertained using the Johns Hopkins ACG System.

Figure 25. Percent of the top 20 chronic disease categories, with highest degree centrality, in common between pairs of select co-occurrence measures in networks limited to strongest 200 statistically significant pairwise associations.



Note: chronic diseases were ascertained using the Johns Hopkins ACG System.

Figure 26. Percent of the top 20 chronic disease categories, with highest degree centrality, in common between pairs of select co-occurrence measures in networks limited to strongest 50 percent (*n*=3,922) of all statistically significant pairwise associations.



Note: chronic diseases were ascertained using the Johns Hopkins ACG System.

Table 18. Top 20 chronic disease categories with highest degree centrality (i.e., most co-occurrence relationships) in pairwise networks limited to the strongest 200 co-occurrence relationships as measured using relative risk, phi, and joint prevalence.

	Relative risk	Phi	Joint prevalence
1	Aortic aneurysm	Hypertension	Hypertension
2	Congestive heart failure	Peripheral vascular disease	Type 2 diabetes
3	Peripheral vascular disease	Congestive heart failure	Degenerative joint disease
4	ESRD	Ischemic heart disease (excluding AMI)	Disorders of lipid metabolism
5	Vesicoureteral reflux	Cardiac arrhythmia	Depression
6	Cerebral palsy	Type 2 diabetes	Ischemic heart disease (excluding AMI)
7	Cardiomyopathy	Degenerative joint disease	Asthma
8	Renal disorders, other	Renal disorders, other	Cardiac arrhythmia
9	Personality disorders	Cataract, aphakia	Hypothyroidism
10	Acute myocardial infarction	Chronic ulcer of the skin	Cataract, aphakia
11	Quadriplegia and paraplegia	Emphysema, chronic bronchitis, COPD	Congestive heart failure
12	Cardiac arrest, shock	Dementia	Obesity
13	Acute respiratory failure	Cerebrovascular disease	Emphysema, chronic bronchitis, COPD
14	Cardiovascular signs and symptoms	Generalized atherosclerosis	Anxiety, neuroses
15	Malignant neoplasms, liver and biliary tract	Chronic renal failure	Dementia
16	Dementia	Depression	Glaucoma
17	Chronic renal failure	Neurologic disorders, other	Sleep apnea
18	Malignant neoplasms, stomach	Cardiovascular disorders, other	Other endocrine disorders
19	Urinary symptoms	Diabetic retinopathy	Deficiency anemias
20	Hypertension	Disorders of lipid metabolism	Cerebrovascular disease

AMI = acute myocardial infarction, COPD = chronic obstructive pulmonary disease, ESRD = end-stage renal disease.
Chronic diseases were ascertained using the Johns Hopkins ACG System.

Table 19. Comparison of the top 20 chronic disease categories with highest degree centrality (i.e., most co-occurrence relationships) between a pairwise network, and a hypergraph that included both pairwise and triad associations.

	Pairwise network	Hypergraph
1	Other endocrine disorders	Hypertension
2	Depression	Degenerative joint disease
3	Major depression	Depression
4	Sleep apnea	Type 2 diabetes
5	Asthma	Ischemic heart disease (excluding AMI)
6	Obesity	Cardiac arrhythmia
7	Cardiovascular disorders, other	Disorders of lipid metabolism
8	Anxiety, neuroses	Emphysema, chronic bronchitis, COPD
9	Emphysema, chronic bronchitis, COPD	Congestive heart failure
10	Respiratory disorders, other	Hypothyroidism
11	Degenerative joint disease	Obesity
12	Hypertension	Cataract, aphakia
13	Musculoskeletal disorders, other	Asthma
14	Neurologic disorders, other	Anxiety, neuroses
15	Autoimmune and connective tissue diseases	Cerebrovascular disease
16	Cardiac arrhythmia	Renal disorders, other
17	Dementia	Sleep apnea
18	Type 2 diabetes	Peripheral vascular disease
19	Hypothyroidism	Other endocrine disorders
20	Deafness, hearing loss	Dementia

AMI = acute myocardial infarction, COPD = chronic obstructive pulmonary disease.
Chronic diseases were ascertained using the Johns Hopkins ACG System.

4.9 Sex-stratified Network Complexity

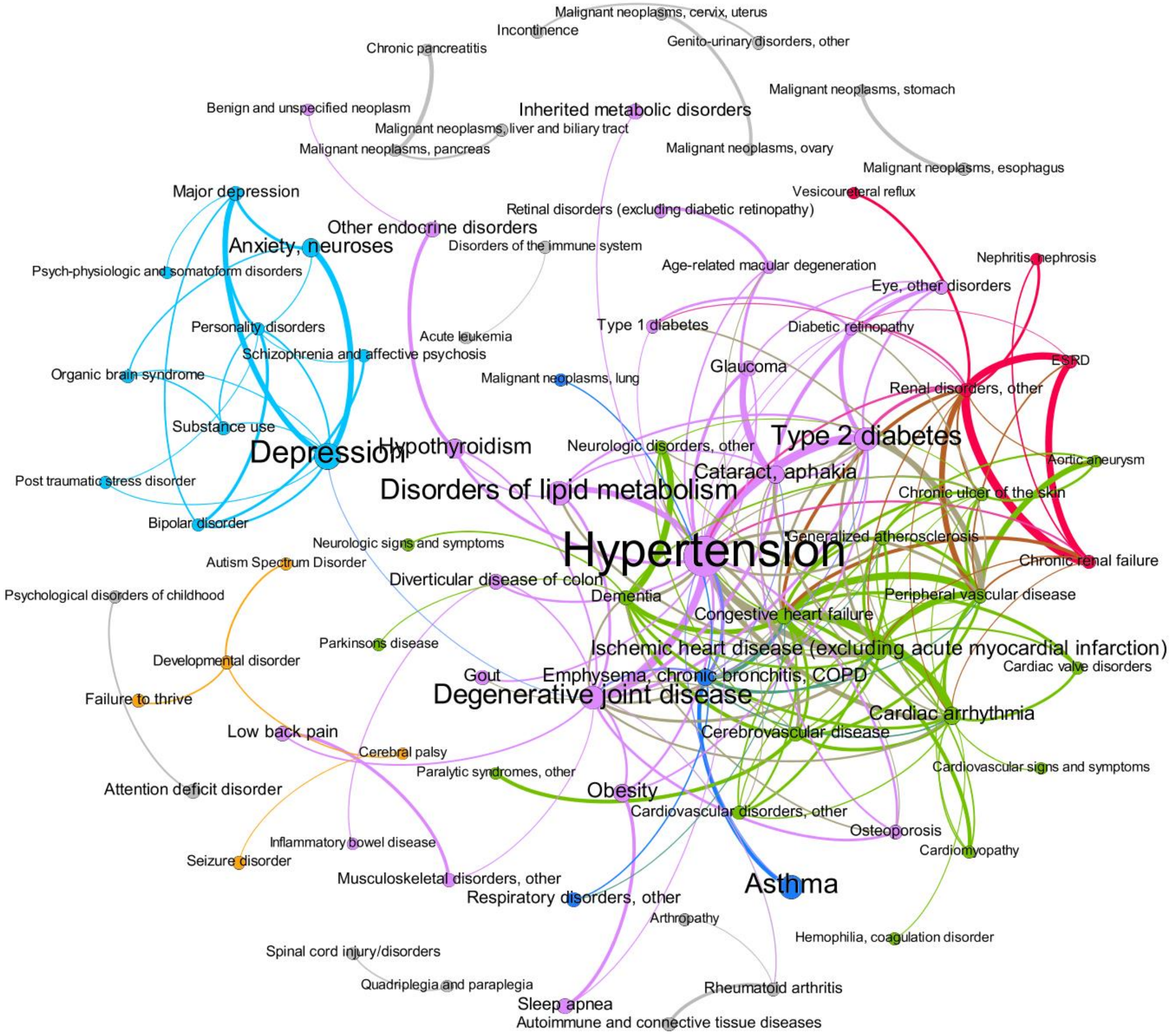
When including all statistically significant associations and measuring disease status using the ACG System, the male and female disease networks had similar network density: 0.52 for the female disease network compared with an estimated density of 0.51 for the male network. Network density was also similar when the Elixhauser index was used for disease ascertainment, with the male network estimated to have slightly higher density (0.96) compared with the female network (0.94). Male and female disease networks were also similar in density when the number of included associations was reduced to the strongest 50% and the strongest 200 of all statistically significant pairwise associations (Table 20, Figure 27, Figure 28).

Table 20. Global properties pairwise networks constructed with select co-occurrence measures and limited to the strongest 50 percent and the strongest 200 of all statistically significant associations, stratified by sex.

Disease co-occurrence inclusion criteria	Association measure	Female			Male		
		N nodes	N edges	Density	N nodes	N edges	Density
Top 50%	Lift	160	3,279	0.26	158	3,134	0.25
	Relative risk	160	3,279	0.26	158	3,134	0.25
	Phi	157	3,279	0.27	152	3,134	0.27
	Jaccard	142	3,279	0.33	135	3,134	0.35
	Cosine	155	3,279	0.28	147	3,134	0.29
	Kulczynski	160	3,279	0.26	158	3,134	0.25
	Joint prevalence	144	3,279	0.32	141	3,134	0.32
Top 200	Lift	101	200	0.04	106	200	0.04
	Relative risk	104	200	0.04	112	200	0.03
	Phi	83	200	0.06	83	200	0.06
	Jaccard	70	200	0.08	72	200	0.08
	Cosine	72	200	0.08	74	200	0.07
	Kulczynski	117	200	0.03	119	200	0.03
	Joint prevalence	56	200	0.13	53	200	0.15

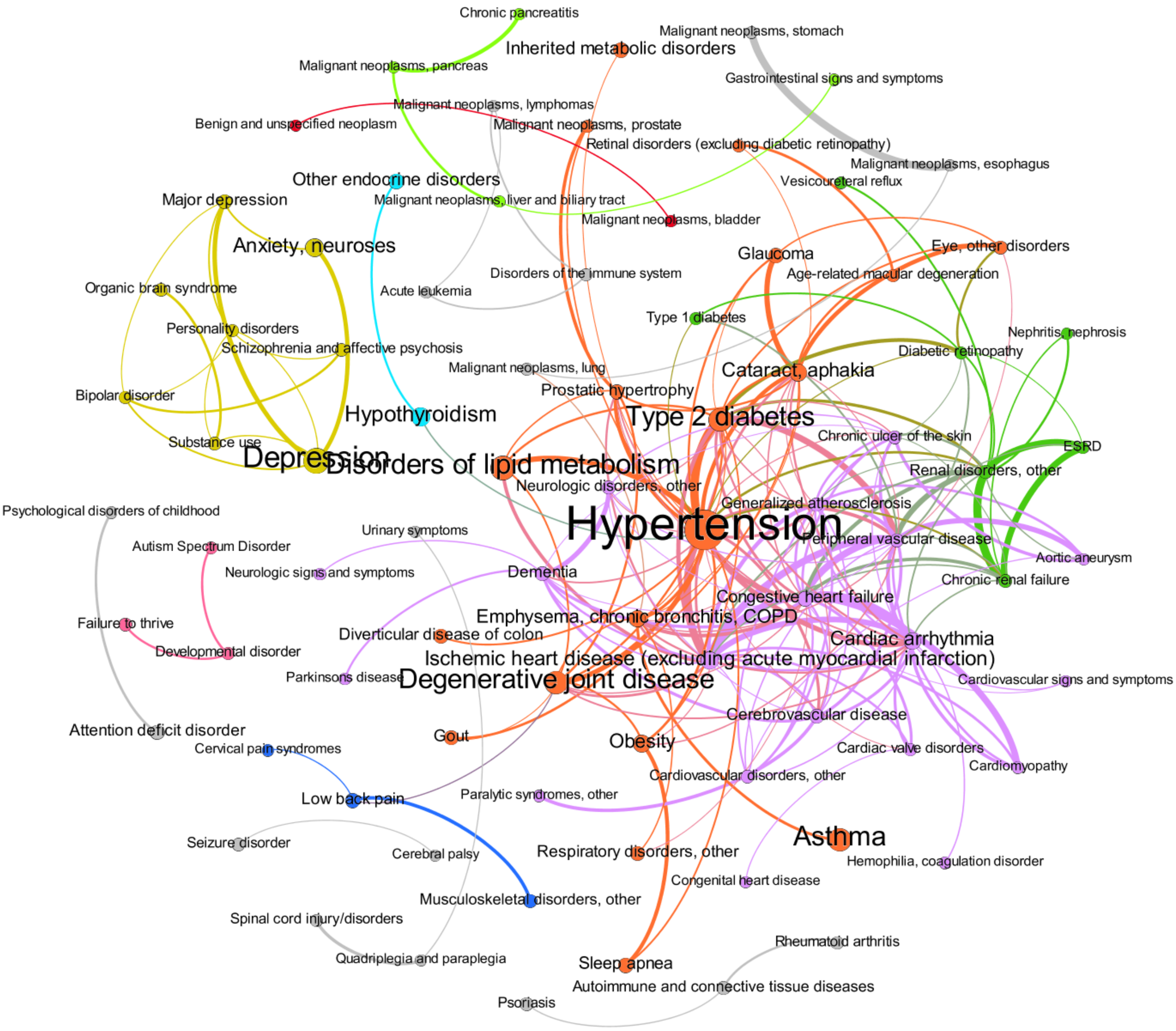
Note: Chronic diseases were ascertained using the Johns Hopkins ACG System.

Figure 27. Female pairwise chronic disease network constructed from the strongest 200 statistically significant associations, with diseases ascertained using the Johns Hopkins ACG System and co-occurrence relationships measured using phi.


Malignant neoplasms, cervix, uterus
Incontinence
Genito-urinary disorders, other
Chronic pancreatitis
Malignant neoplasms, stomach
Benign and unspecified neoplasm
Inherited metabolic disorders
Malignant neoplasms, liver and biliary tract
Malignant neoplasms, pancreas
Malignant neoplasms, ovary
Malignant neoplasms, esophagus
Major depression
Vesicoureteral reflux
Retinal disorders (excluding diabetic retinopathy)
Other endocrine disorders
Anxiety, neuroses
Disorders of the immune system
Nephritis, nephrosis
Psych-physiologic and somatoform disorders
Age-related macular degeneration
Eye, other disorders
Type 1 diabetes
Personality disorders
Diabetic retinopathy
Acute leukemia
Schizophrenia and affective psychosis
ESRD
Organic brain syndrome
Malignant neoplasms, lung
Glaucoma
Renal disorders, other
Substance use
Type 2 diabetes
Depression
Hypothyroidism
Neurologic disorders, other
Cataract, aphakia
Aortic aneurysm
Post traumatic stress disorder
Disorders of lipid metabolism
Chronic ulcer of the skin
Bipolar disorder
Generalized atherosclerosis
Neurologic signs and symptoms
Hypertension
Autism Spectrum Disorder
Chronic renal failure
Diverticular disease of colon
Psychological disorders of childhood
Dementia
Peripheral vascular disease
Congestive heart failure
Parkinsons disease
Ischemic heart disease (excluding acute myocardial infarction)
Developmental disorder
Cardiac valve disorders
Gout
Emphysema, chronic bronchitis, COPD
Failure to thrive
Degenerative joint disease
Cardiac arrhythmia
Low back pain
Cerebrovascular disease
Cerebral palsy
Paralytic syndromes, other
Cardiovascular signs and symptoms
Attention deficit disorder
Obesity
Cardiovascular disorders, other
Osteoporosis
Inflammatory bowel disease
Cardiomyopathy
Seizure disorder
Musculoskeletal disorders, other
Asthma
Respiratory disorders, other
Arthropathy
Hemophilia, coagulation disorder
Spinal cord injury/disorders
Quadriplegia and paraplegia
Rheumatoid arthritis
Sleep apnea
Autoimmune and connective tissue diseases

Note: Node diameter and font size are proportional to disease prevalence, edge weight (thickness) is proportional to effect size, and node and edge colour indicate community structure (i.e., disease clusters).
COPD = chronic obstructive pulmonary disease, ESRD = end-stage renal disease.

Figure 28. Male pairwise chronic disease network constructed from the strongest 200 statistically significant associations, with diseases ascertained using the Johns Hopkins ACG System and co-occurrence relationships measured using phi.


Chronic pancreatitis
Malignant neoplasms, stomach
Inherited metabolic disorders
Malignant neoplasms, pancreas
Gastrointestinal signs and symptoms
Malignant neoplasms, lymphomas
Benign and unspecified neoplasm
Malignant neoplasms, prostate
Retinal disorders (excluding diabetic retinopathy)
Malignant neoplasms, esophagus
Other endocrine disorders
Vesicoureteral reflux
Major depression
Malignant neoplasms, liver and biliary tract
Malignant neoplasms, bladder
Anxiety, neuroses
Glaucoma
Eye, other disorders
Organic brain syndrome
Disorders of the immune system
Age-related macular degeneration
Acute leukemia
Type 1 diabetes
Nephritis, nephrosis
Personality disorders
Schizophrenia and affective psychosis
Diabetic retinopathy
Malignant neoplasms, lung
Cataract, aphakia
Bipolar disorder
Prostatic hypertrophy
Hypothyroidism
Type 2 diabetes
Chronic ulcer of the skin
Substance use
ESRD
Depression
Disorders of lipid metabolism
Renal disorders, other
Neurologic disorders, other
Psychological disorders of childhood
Generalized atherosclerosis
Urinary symptoms
Hypertension
Autism Spectrum Disorder
Peripheral vascular disease
Aortic aneurysm
Neurologic signs and symptoms
Dementia
Chronic renal failure
Congestive heart failure
Failure to thrive
Emphysema, chronic bronchitis, COPD
Diverticular disease of colon
Cardiac arrhythmia
Developmental disorder
Ischemic heart disease (excluding acute myocardial infarction)
Parkinsons disease
Degenerative joint disease
Cardiovascular signs and symptoms
Cerebrovascular disease
Attention deficit disorder
Gout
Cervical pain syndromes
Obesity
Cardiac valve disorders
Cardiomyopathy
Cardiovascular disorders, other
Low back pain
Paralytic syndromes, other
Asthma
Seizure disorder
Cerebral palsy
Respiratory disorders, other
Hemophilia, coagulation disorder
Congenital heart disease
Musculoskeletal disorders, other
Spinal cord injury/disorders
Rheumatoid arthritis
Quadriplegia and paraplegia
Sleep apnea
Autoimmune and connective tissue diseases
Psoriasis

Note: Node diameter and font size are proportional to disease prevalence, edge weight (thickness) is proportional to effect size, and node and edge colour indicate community structure (i.e., disease clusters).
COPD = chronic obstructive pulmonary disease, ESRD = end-stage renal disease.

Chapter 5: Discussion

Measuring disease co-occurrence is essential when constructing chronic disease networks to determine the connecting links between disease nodes and the strengths of these co-occurrence relationships. Different association measures highlight different co-occurrence relationships, in terms of disease prevalence, based on which relationships are assigned higher association estimates. In weighted disease networks where effect size estimates are used as edge weights, differences in co-occurrence measurement will influence community detection algorithms and node centrality measures that use edge weights in their calculations. Unweighted measures such as network density and degree centrality will not be affected by choice of co-occurrence measure unless network links are excluded based on effect size cut-offs. When limiting the number of edges in a network by effect size, to produce a visually interpretable diagram, the choice of co-occurrence measure can have a significant impact on network structure and network analysis inferences. Evaluating the accuracy or validity of a network requires a ground truth against which to compare network structure. Since there is no ground truth for a chronic disease co-occurrence network, this study performed a descriptive analysis to highlight the impact that co-occurrence measurement has on network analysis.

5.1 Summary of Key Findings

This study showed the majority of the highest associations measured using lift and relative risk pertained to co-occurrence relationships between pairs of low prevalent conditions. In contrast, the strongest associations in the joint prevalence network included highly prevalent conditions, while the Kulczynski measure emphasized relationships between high and low prevalent diseases. Phi, Jaccard, and cosine emphasized associations with moderately prevalent conditions. Comparing Jaccard and cosine, Jaccard tended to prefer co-occurrence relationships between diseases of similar prevalence, while cosine assigned slightly less emphasis to events of similar frequency. Distinctions in the prevalence difference distributions resulted in significant dissimilarities in community detection and centrality analysis, two of the main components of a network analysis. However, choice of co-occurrence measure was not found to considerably affect comparisons of network density between male and female disease networks.

Many chronic diseases co-occur in groups of three or more and limiting network analyses to pairwise associations does not adequately depict the real-world complexity of multimoribidty.

Higher-order disease associations can be extracted using association rule mining and modeled using hypergraphs. Parallel Aggregated Ordered Hypergraph (PAOH) diagrams, alternative hypergraph visualizations, have higher visual interpretability than traditional hypergraph diagrams while depicting a larger number of associations. When comparing hypergraph-based disease networks with their respective pairwise networks constructed using the same co-occurrence measure, significant differences were observed in terms of their agreement on the most central nodes and the pairwise relationships held in common.

5.1 Context of Study Findings within Literature

The results from the current study concur with the results of the study by Hidalgo et al., who compared disease co-occurrence networks constructed using RR and ϕ and found the network constructed with RR to have a greater number of low prevalence conditions and the ϕ-based network to be characterized by more prevalent conditions.[11] In addition to describing network edges by disease prevalence, the current study also showed the impact that co-occurrence measurement has on community structure, node centrality, and subgroup comparisons—items not discussed previously in literature. Along with contrasting RR and ϕ, this study also compared disease networks constructed using lift, a measure commonly used in conjunction with association rule mining, and null-invariant measures suggested for use with sparse datasets such as disease status matrices. The differences amongst the null-invariant measures observed in the current study agree with Wu et al., who described the preference of Jaccard for relationships between events of similar frequency, Kulczynski for relationships between frequent and rare events, and cosine as being situated between these two in terms of the relationships that receive the highest association estimates.[83]

Several previous network analyses identified associations amongst combinations of three or more diseases, but limited network visualizations to pairwise graphs by flattening the higher-order associations into their respective binary relationships; this results in a loss of information.[20–30] The current study went a step further and demonstrated how the additional information present in multi-way disease associations could be modeled and analyzed using hypergraphs; future research involving higher-order disease co-occurrence relationships could benefit from visualizations that depict complex multimorbidity relationships.

5.2 Study Strengths

The current study has a number of strengths. Extracting diagnoses from both hospital and physician data aids in providing a comprehensive picture of chronic disease patterns in the Manitoba population. Furthermore, the administrative health data used in this study had excellent population coverage since the data are based on a single public insurer that effectively captures healthcare system encounters for all Manitoba residents, with few exceptions—resulting in excellent generalizability of the observed chronic disease patterns at the population level. Utilizing 5-digit ICD diagnostic codes minimized misclassification errors and allowed for the definition of certain disease categories that cannot be distinguished from one another when only using 3-digit codes.

The large number of chronic disease categories under analysis facilitated the examination of many potentially interesting disease patterns that are obscured when using a more limited number of disease categories based on a comorbidity index. Using a relatively large number of chronic disease categories is beneficial for hypothesis generation. By reporting results separately for different network sizes (i.e., when including the top 200 associations, top 50 percent of associations, and all associations) and stratifying by disease ascertainment method (i.e., Elixhauser comorbidity index, or Johns Hopkins ACG System), the results from this study are applicable to many different types of network analyses.

Besides exploring the effect that co-occurrence measurement has on disease networks and demonstrating the use of hypergraphs, this study also provides insight into patterns of co-occurring chronic disease at the population level and is available for further exploration by chronic disease researchers or policy makers. Finally, the included literature review adds to the work done by Brunson and Laubenbacher[84] to summarize the methodology of published disease network analyses and link together this body of literature.

5.3 Study Limitations

Despite the strengths of this study, there are some limitations. The true distribution of chronic disease in the underlying population can differ significantly from disease patterns observed within administrative claims data, where disease status accuracy is dependent upon individuals coming into contact with the healthcare system and upon billing codes accurately portraying patient health profiles. Factors leading to non-representative reporting of disease patterns within this retrospective claims-based study include differential healthcare utilization patterns, observation period limitations, “rule out” diagnostic practices, and diagnostic coding errors.

Because diseases were defined through contact with the healthcare system, disease information may have been inadequately captured for individuals with limited access to healthcare services or conditions for which individuals are less likely to seek treatment. Resulting bias would have been incurred if disease patterns were significantly different for the individuals that are less likely to seek treatment, in comparison to the general population. Consequently, there will be missing links or underestimated edge weights for relationships involving underreported health conditions within the structure of the disease co-occurrence networks. To increase diagnostic precision, this study was constrained to the 4-year period of time when physician billing claims were coded with 5-digit ICD codes; but in doing so this study did not capture diagnoses that were only recorded in earlier time periods. This reduced observation period may have resulted in understating co-occurrence for less prevalent conditions or conditions that are infrequently documented in billing claims.

All diagnoses observed during the 4-year study period for a specific individual were treated as persisting during the entire time period and assumed to co-occur with one another. This may have resulted in overstating certain co-occurrence relationships, since diseases that may have been in remission were still considered as co-occurring with other conditions after the point of remission. Diagnoses that did not map to any of the 167 EDC categories or the 31 Elixhauser comorbidities were also excluded from the analysis, resulting in missing network links between network nodes and any omitted chronic condition categories. Due to the relatively large number of disease categories under consideration, it was not feasible to use complex case definitions to ascertain disease status based on diagnosis code counts. Simplified case definitions based on single diagnosis codes were used to mark disease status and misclassification may have occurred due to diagnostic coding errors, or the presence of “rule out” diagnoses when clinicians are working with patients to resolve health concerns—leading to overestimating co-occurrence with conditions overreported within billing claims.

This study contrasted seven co-occurrence measures in the context of a chronic disease network analysis, but it was not feasible to also investigate all other association measures of potential interest to researchers. For the same reason, this analysis limited community detection to a single non-overlapping detection algorithm, centrality analysis to node degree, and network complexity to density measurement. Evaluating other community detection algorithms, and different centrality measures such as eigenvector or betweenness centrality, would provide additional insight into the effect of co-occurrence measurement on network analysis. Descriptive analysis was used to quantify differences in network metrics among networks constructed using different association measures, but statistical significance testing was not used since the research purpose was to describe the overall effect of co-occurrence measurement and the research was not focused on testing hypotheses of differences between individual networks or drawing inferences on the underlying population. Furthermore, software restrictions within the secure data environment, which houses the Manitoba Population Health Research Data Repository, posed challenges for calculating empirical standard error estimates of network measures.

5.4 Applications and Next Steps

The differences observed between disease networks constructed with different association measures suggest researchers should select co-occurrence measures based on the prevalence relationships of greatest interest, and their specific research objectives (e.g., hypothesis generation, data visualization). If researchers are seeking to explore associations between highly prevalent and low prevalent conditions, then Kulczynski may be an appropriate choice based on its tendency to assign high association estimates towards skewed relationships. Whereas, the preference of relative risk and lift make these measures suitable for exploring relationships between pairs of low prevalent conditions. Phi, Jaccard, and cosine are appropriate for analyzing co-occurrence relationships involving moderately prevalent diseases. Joint prevalence has an interpretability advantage over many of the other co-occurrence measures, which may make it more suitable for knowledge translation activities with non-technical audiences, specifically if relationships between the most prevalent conditions are of interest. Knowing the tendencies of different co-occurrence measures will allow researchers to make an informed choice based on their research goals. Although this study highlighted differences when networks were limited to the strongest associations, researchers may instead choose other effect size ranges such as the lowest or intermediate estimates, depending upon on their study objectives.

Software implementations of hypergraph analytic techniques are available for researchers seeking to incorporate knowledge of higher-order associations into a network analysis. Bipartite representations promote the analysis of hypergraphs using standard network analysis software, but converting hypergraphs to bipartite graphs modifies the network structure, which may not be desirable. Current software supports the visualization of hyperedges as standard coloured bounding containers or as vertical lines in the alternative PAOH figure, but hypergraph visualizations may be more difficult to interpret than pairwise network diagrams and analysts should consider which approach is best given their objectives. Further development of hypergraph analytic software will improve the viability of multi-way association analysis and visualization.

Researchers must make several methodological choices when seeking to conduct a network analysis. In addition to choosing a measure of association, researchers must choose from many different community detection techniques, and node centrality and network complexity measures. While this study discusses approaches to choosing an association measure, researchers seeking to conduct an analysis of a disease co-occurrence network will also benefit from additional guidelines on choosing from these other network methods.

Administrative health data is available in all jurisdictions within Canada. The methodology used in the current study can be readily applied to compare population-level chronic disease patterns across the Canadian provinces and territories and within population sub-groups defined by determinants of health.

5.5 Conclusion

Disease co-occurrence measurement has a significant effect on the structure of chronic disease co-occurrence networks and influences which diseases are considered dominant within a population (i.e., node centrality), how disease clusters are defined (i.e., network community structure), and characterizations of disease network complexity. Choice of co-occurrence measure considerably affects our understanding of population-level chronic disease patterns obtained using network analysis. Co-occurrence measures should be selected considering research objectives and the prevalence relationships of greatest interest. Researchers should be cautious when interpreting results from network analyses of co-occurring chronic disease and should conduct sensitivity analyses using different co-occurrence measures. Finally, many chronic diseases co-occur in groups of three or more and these higher-order associations can be effectively visualized and analyzed using hypergraph techniques.

Literature Cited

1. Koné Pefoyo, A. J. *et al.* The increasing burden and complexity of multimorbidity disease epidemiology - Chronic. *BMC Public Health* **15**, 1–11 (2015).

2. Afshar, S., Roderick, P. J., Kowal, P., Dimitrov, B. D. & Hill, A. G. Multimorbidity and the inequalities of global ageing: A cross-sectional study of 28 countries using the World Health Surveys. *BMC Public Health* **15**, 776 (2015).

3. Roberts, K. C., Rao, D. P., Bennett, T. L., Loukine, L. & Jayaraman, G. C. Prevalence and patterns of chronic disease multimorbidity and associated determinants in Canada. *Heal. Promot. chronic Dis. Prev. Canada Res. policy Pract.* **35**, 87–94 (2015).

4. Tsasis, P. & Bains, J. Management of complex chronic disease: facing the challenges in the Canadian health-care system. *Heal. Serv. Manag. Res.* **21**, 228–235 (2008).

5. Moffat, K. & Mercer, S. W. Challenges of managing people with multimorbidity in today's healthcare systems. *BMC Fam. Pract.* **16**, 129 (2015).

6. Thavorn, K. *et al.* Effect of socio-demographic factors on the association between multimorbidity and healthcare costs: A population-based, retrospective cohort study. *BMJ Open* **7**, (2017).

7. Reid, R. *et al.* Conspicuous consumption: Characterizing high users of physician services in one Canadian province. *J. Heal. Serv. Res. Policy* **8**, 215–224 (2003).

8. Chmiel, A., Klimek, P. & Thurner, S. Spreading of diseases through comorbidity networks across life and gender. *New J. Phys.* **16**, 115013 (2014).

9. Duarte, C. W., Lindner, V., Francis, S. A. & Schoormans, D. Visualization of Cancer and Cardiovascular Disease Co-Occurrence With Network Methods. *JCO Clin. cancer informatics* **1**, 1–12 (2017).

10. Hanauer, D. A. & Ramakrishnan, N. Modeling temporal relationships in large scale clinical associations. *J. Am. Med. Informatics Assoc.* **20**, 332–341 (2013).

11. Hidalgo, C. A., Blumm, N., Barabási, A.-L. & Christakis, N. A. A dynamic network approach for the study of human phenotypes. *PLoS Comput. Biol.* **5**, e1000353 (2009).

12. Davis, D. A. & Chawla, N. V. Exploring and exploiting disease interactions from multi-relational gene and phenotype networks. *PLoS One* **6**, e22670 (2011).

13. Jeong, E., Ko, K., Oh, S. & Han, H. W. Network-based analysis of diagnosis progression patterns using claims data. *Sci. Rep.* **7**, 1–12 (2017).

14. Jiang, Y., Ma, S., Shia, B. C. & Lee, T. S. An epidemiological human disease network derived from disease Co-occurrence in Taiwan. *Sci. Rep.* **8**, 1–12 (2018).

15. Kalgotra, P., Sharda, R. & Croff, J. M. Examining health disparities by gender: A multimorbidity network analysis of electronic medical record. *Int. J. Med. Inform.* **108**, 22–28 (2017).

16. Khan, A., Uddin, S. & Srinivasan, U. Comorbidity network for chronic disease: A novel

approach to understand type 2 diabetes progression. *Int. J. Med. Inform.* **115**, 1–9 (2018).

17. Kim, J. H. *et al.* Network analysis of human diseases using Korean nationwide claims data. *J. Biomed. Inform.* **61**, 276–282 (2016).

18. Lai, Y. H. A network approach for the comorbidities of HIV/AIDS in Taiwan. in *Technology and Health Care* vol. 24 S377–S383 (IOS Press, 2016).

19. Moni, M. A. & Liò, P. Network-based analysis of comorbidities risk during an infection: SARS and HIV case studies. *BMC Bioinformatics* **15**, 333 (2014).

20. Schäfer, I. *et al.* Reducing complexity: A visualisation of multimorbidity by combining disease clusters and triads. *BMC Public Health* **14**, 1285 (2014).

21. Tai, Y.-M. & Chiu, H.-W. Comorbidity study of ADHD: Applying association rule mining (ARM) to National Health Insurance Database of Taiwan. *Int. J. Med. Inform.* **78**, e75–e83 (2009).

22. Zheng, C. & Xu, R. The Alzheimer's comorbidity phenome: mining from a large patient database and phenome-driven genetics prediction. *JAMIA Open* **2**, 131–138 (2019).

23. Chen, Y. & Xu, R. Mining Cancer-Specific Disease Comorbidities from a Large Observational Health Database. *Cancer Inform.* **13s1**, CIN.S13893 (2014).

24. Chen, Y., Li, L. & Xu, R. Disease Comorbidity Network Guides the Detection of Molecular Evidence for the Link Between Colorectal Cancer and Obesity. *AMIA Jt. Summits Transl. Sci. proceedings. AMIA Jt. Summits Transl. Sci.* **2015**, 201–6 (2015).

25. Held, F. P. *et al.* Association rules analysis of comorbidity and multimorbidity: The Concord Health and Aging in Men Project. *Journals Gerontol. Ser. A Biomed. Sci. Med. Sci.* **71**, 625–631 (2015).

26. Hernández, B., Reilly, R. B. & Kenny, R. A. Investigation of multimorbidity and prevalent disease combinations in older Irish adults using network analysis and association rules. *Sci. Rep.* **9**, 1–12 (2019).

27. Kim, L. & Myoung, S. Comorbidity Study of Attention-deficit Hyperactivity Disorder (ADHD) in Children: Applying Association Rule Mining (ARM) to Korean National Health Insurance Data. *Iran. J. Public Health* **47**, 481–488 (2018).

28. Shin, A. M. *et al.* Diagnostic Analysis of Patients with Essential Hypertension Using Association Rule Mining. *Healthc. Inform. Res.* **16**, 77 (2010).

29. Zemedikun, D. T., Gray, L. J., Khunti, K., Davies, M. J. & Dhalwani, N. N. Patterns of multimorbidity in middle-aged and older adults: an analysis of the UK Biobank data. in *Mayo Clinic Proceedings* vol. 93 857–866 (Elsevier, 2018).

30. Zheng, C. & Xu, R. Large-scale mining disease comorbidity relationships from post-market drug adverse events surveillance data. *BMC Bioinformatics* **19**, 500 (2018).

31. Prados-Torres, A., Calderón-Larrañaga, A., Hancco-Saavedra, J., Poblador-Plou, B. & van den Akker, M. Multimorbidity patterns: a systematic review. *J. Clin. Epidemiol.* **67**, 254–266 (2014).

32. Ng, S. K., Tawiah, R., Sawyer, M. & Scuffham, P. Patterns of multimorbid health conditions: a systematic review of analytical methods and comparison analysis. *Int. J. Epidemiol.* **47**, 1687–1704 (2018).

33. Divo, M. J. *et al.* COPD comorbidities network. *Eur. Respir. J.* **46**, 640–650 (2015).

34. Wu, T., Chen, Y. & Han, J. Association mining in large databases: A re-examination of its measures. in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* vol. 4702 LNAI 621–628 (Springer Verlag, 2007).

35. Kim, S., Barsky, M. & Han, J. Efficient mining of top correlated patterns based on null-invariant measures. in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* vol. 6912 LNAI 177–192 (Springer, Berlin, Heidelberg, 2011).

36. Wu, T., Chen, Y. & Han, J. Re-examination of interestingness measures in pattern mining: A unified framework. *Data Min. Knowl. Discov.* **21**, 371–397 (2010).

37. Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J. R. Stat. Soc. Ser. B* **57**, 289–300 (1995).

38. Celli, R., Divo, M., Colunga, M., Celli, B. & Mitchell-Richards, K. Network analysis of autopsy diagnoses: Insights into the "cause of death" from unbiased disease clustering. *J. Pathol. Inform.* **9**, 35 (2018).

39. Cao, H., Hripcsak, G. & Markatou, M. A statistical methodology for analyzing co-occurrence data from a large sample. *J. Biomed. Inform.* **40**, 343–352 (2007).

40. Agrawal, R. & Srikant, R. Fast algorithms for mining association rules. in *Proceedings of the 20th International Conference on Very Large Data Bases* vol. 1215 487–499 (1994).

41. Tan, P.-N., Steinbach, M. & Kumar, V. *Introduction to Data Mining*. (Pearson Addison Wesley, 2005).

42. Koh, Y. S. & Rountree, N. *Rare association rule mining and knowledge discovery : technologies for infrequent and critical event detection*. (Information Science Reference, 2010).

43. Brijs, T., Vanhoof, K. & Wets, G. Defining interestigness for association rules. (2003).

44. Han, J., Kamber, M. & Pei, J. *Data mining : concepts and techniques*. (Elsevier Science, 2011).

45. McNicholas, P. D., Murphy, T. B. & O'Regan, M. Standardising the lift of an association rule. *Comput. Stat. Data Anal.* **52**, 4712–4721 (2008).

46. Hatahira, H. *et al.* Analysis of fall-related adverse events among older adults using the Japanese Adverse Drug Event Report (JADER) database. *J. Pharm. Heal. Care Sci.* **4**, 32 (2018).

47. Yildirim, P. Association patterns in open data to explore ciprofloxacin adverse events. *Appl. Clin. Inform.* **6**, 728–747 (2015).

48. Shen, C.-C., Hu, L.-Y. & Hu, Y.-H. Comorbidity study of borderline personality disorder: applying association rule mining to the Taiwan national health insurance research database. *BMC Med. Inform. Decis. Mak.* **17**, 8 (2017).

49. Han, J., Pei, J. & Yin, Y. Mining frequent patterns without candidate generation. in *Proceedings of the 2000 ACM SIGMOD International Conference on Management of Data* vol. 29 1–12 (SIGMOD '00, 2000).

50. Zaki, M. J. Scalable algorithms for association mining. *IEEE Trans. Knowl. Data Eng.* **12**, 372–390 (2000).

51. Ho, V. P., Schiltz, N. K., Reimer, A. P., Madigan, E. A. & Koroukian, S. M. High-Risk Comorbidity Combinations in Older Patients Undergoing Emergency General Surgery. *J. Am. Geriatr. Soc.* **67**, 503–510 (2019).

52. M. Kang'ethe, S. & W. Wagacha, P. Extracting Diagnosis Patterns in Electronic Medical Records using Association Rule Mining. *Int. J. Comput. Appl.* **108**, 19–26 (2014).

53. Kim, H. S., Shin, A. M., Kim, M. K. & Kim, Y. N. Comorbidity study on type 2 diabetes mellitus using data mining. *Korean J. Intern. Med.* **27**, 197–202 (2012).

54. Madlock-Brown, C. & Reynolds, R. B. Identifying obesity-related multimorbidity combinations in the United States. *Clin. Obes.* **9**, e12336 (2019).

55. Nassar, Y. & Richter, S. Gastroparesis in Non-Diabetics: Associated Conditions and Possible Risk Factors. *Gastroenterol. Res.* **11**, 340–345 (2018).

56. Peng, M. *et al.* Exploration of association rule mining for coding consistency and completeness assessment in inpatient administrative health data. *J. Biomed. Inform.* **79**, 41–47 (2018).

57. Valent, F., Tillati, S. & Zanier, L. Prevalence and comorbidities of known diabetes in northeastern Italy. *J. Diabetes Investig.* **4**, 355–360 (2013).

58. Wang, C.-H., Lee, T.-Y., Hui, K.-C. & Chung, M.-H. Mental disorders and medical comorbidities: Association rule mining approach. *Perspect. Psychiatr. Care* **55**, 517–526 (2019).

59. Yao, S.-S. *et al.* Prevalence and Patterns of Multimorbidity in a Nationally Representative Sample of Older Chinese: Results From the China Health and Retirement Longitudinal Study. *Journals Gerontol. Ser. A* (2019) doi:10.1093/gerona/glz185.

60. Estrada, E. & Rodriguez-Velazquez, J. A. Complex networks as hypergraphs. *arXiv Prepr. physics/0505137* (2005).

61. Benson, A. R. Three hypergraph eigenvector centralities. *SIAM J. Math. Data Sci.* **1**, 293–312 (2019).

62. Fotouhi, B., Momeni, N., Riolo, M. A. & Buckeridge, D. L. Statistical methods for constructing disease comorbidity networks from longitudinal inpatient data. *Appl. Netw. Sci.* **3**, 46 (2018).

63. Doulis, M. Robustness of disease association measures and network structure in an

evolving synthetic cohort. (2016).

64. Valdivia, P., Buono, P., Plaisant, C., Dufournaud, N. & Fekete, J.-D. Analyzing Dynamic Hypergraphs with Parallel Aggregated Ordered Hypergraph Visualization. *IEEE Trans. Vis. Comput. Graph.* (2019).

65. Hypergraph. *Encyclopedia of Mathematics* https://encyclopediaofmath.org/index.php?title=Hypergraph.

66. Niu, Y.-W., Qu, C.-Q., Wang, G.-H. & Yan, G.-Y. RWHMDA: Random Walk on Hypergraph for Microbe-Disease Association Prediction. *Front. Microbiol.* **10**, 1578 (2019).

67. Jie, B., Wee, C. Y., Shen, D. & Zhang, D. Hyper-connectivity of functional networks for brain disease diagnosis. *Med. Image Anal.* **32**, 84–100 (2016).

68. Mukhopadhyay, S., Palakal, M. & Maddu, K. Multi-way association extraction and visualization from biological text documents using hyper-graphs: applications to genetic association studies for diseases. *Artif. Intell. Med.* **49**, 145–154 (2010).

69. Belyi, E. *et al.* Combining association rule mining and network analysis for pharmacosurveillance. *J. Supercomput.* **72**, 2014–2034 (2016).

70. Feely, A., Lix, L. M. & Reimer, K. Estimating multimorbidity prevalence with the Canadian Chronic Disease Surveillance System. *Health Promotion and Chronic Disease Prevention in Canada* vol. 37 215–222 (2017).

71. Majumdar, U. B. *et al.* Multiple chronic conditions at a major urban health system: A retrospective cross-sectional analysis of frequencies, costs and comorbidity patterns. *BMJ Open* **9**, (2019).

72. Elixhauser, A., Steiner, C., Harris, D. R. & Coffey, R. M. Comorbidity Measures for Use with Administrative Data. *Med. Care* **36**, 8–27 (1998).

73. The Johns Hopkins University Bloomberg School of Public Health. *The Johns Hopkins ACG® System Version 12.0 User Documentation*. (2019).

74. Tan, P. N., Kumar, V. & Srivastava, J. Selecting the right objective measure for association analysis. in *Information Systems* vol. 29 293–313 (Pergamon, 2004).

75. Brin, S., Motwani, R., Ullman, J. D. & Tsur, S. Dynamic Itemset Counting and Implication Rules for Market Basket Data. *SIGMOD Rec. (ACM Spec. Interes. Gr. Manag. Data)* **26**, 255–264 (1997).

76. Aschengrau, A. & Seage, G. R. *Essentials of Epidemiology in Public Health*. (Jones & Bartlett Learning, 2014).

77. Hahsler, M., Grün, B. & Hornik, K. arules - A Computational Environment for Mining Association Rules and Frequent Item Sets. *J. Stat. Software; Vol 1, Issue 15* (2005) doi:10.18637/jss.v014.i15.

78. Blondel, V. D., Guillaume, J. L., Lambiotte, R. & Lefebvre, E. Fast unfolding of communities in large networks. *J. Stat. Mech. Theory Exp.* **2008**, (2008).

79. Kamiński, B., Poulin, V., Prałat, P., Szufel, P. & Théberge, F. Clustering via hypergraph modularity. *PLoS One* **14**, (2019).

80. Fruchterman, T. M. J. & Reingold, E. M. Graph drawing by force-directed placement. *Softw. Pract. Exp.* **21**, 1129–1164 (1991).

81. Joslyn, C. *et al.* High Performance Hypergraph Analytics of Domain Name System Relationships. in *Hawaii International Conference on System Sciences 2019 Symposium on Cybersecurity Big Data Analytics* (2019).

82. Hubert, L. & Arabie, P. Comparing partitions. *J. Classif.* **2**, 193–218 (1985).

83. Wu, T., Chen, Y. & Han, J. Re-examination of interestingness measures in pattern mining: A unified framework. *Data Min. Knowl. Discov.* **21**, 371–397 (2010).

84. Brunson, J. C. & Laubenbacher, R. C. Applications of network analysis to routinely collected health care data: A systematic review. *Journal of the American Medical Informatics Association* vol. 25 210–221 (2018).

Appendix A. Diagnosis codes for chronic disease ascertainment, based on the Elixhauser comorbidity index, in the Medical Services and Hospital Abstracts databases.

Chronic disease category	Medical Services ICD-9-CM diagnosis codes	Hospital Abstracts ICD-10-CA diagnosis codes
Alcohol abuse	265.2, 291.1, 291.2, 291.3, 291.5, 291.8, 291.9, 303.0, 303.9, 305.0, 357.5, 425.5, 535.3, 571.0, 571.1, 571.2, 571.3, 980, V11.3	F10, E52, G62.1, I42.6, K29.2, K70.0, K70.3, K70.9, T51, Z50.2, Z71.4, Z72.1
Blood loss anemia	280.0	D50.0
Cardiac arrhythmia	426.0, 426.13, 426.7, 426.9, 426.10, 426.12, 427.0, 427.1, 427.2, 427.3, 427.4, 427.6, 427.8, 427.9, 785.0, 996.01, 996.04, V45.0, V53.3	I44.1, I44.2, I44.3, I45.6, I45.9, I47, I48, I49, R00.0, R00.1, R00.8, T82.1, Z45.0, Z95.0
Chronic pulmonary disease	416.8, 416.9, 490, 491, 492, 493, 494, 495, 496, 500, 501, 502, 503, 504, 505, 506.4, 508.1, 508.8	I27.8, I27.9, J40, J41, J42, J43, J44, J45, J46, J47, J60, J61, J62, J63, J64, J65, J66, J67, J68.4, J70.1, J70.3
Coagulopathy	286, 287.1, 287.3, 287.4, 287.5	D65, D66, D67, D68, D69.1, D69.3, D69.4, D69.5, D69.6
Congestive heart failure	398.91, 402.01, 402.11, 402.91, 404.01, 404.03, 404.11, 404.13, 404.91, 404.93, 425.4, 425.5, 425.7, 425.8, 425.9, 428	I09.9, I11.0, I13.0, I13.2, I25.5, I42.0, I42.5, I42.6, I42.7, I42.8, I42.9, I43, I50, P29.0
Deficiency anemia	280.1, 280.8, 280.9, 281	D50.8, D50.9, D51, D52, D53
Depression	296.2, 296.3, 296.5, 300.4, 309, 311	F20.4, F31.3, F31.4, F31.5, F32, F33, F34.1, F41.2, F43.2
Diabetes, with complications	250.4, 250.5, 250.6, 250.7, 250.8, 250.9	E10.2, E10.3, E10.4, E10.5, E10.6, E10.7, E10.8, E11.2, E11.3, E11.4, E11.5, E11.6, E11.7, E11.8, E12.2, E12.3, E12.4, E12.5, E12.6, E12.7, E12.8 , E13.2, E13.3, E13.4, E13.5, E13.6, E13.7, E13.8, E14.2, E14.3, E14.4, E14.5, E14.6, E14.7, E14.8
Diabetes, without complications	250.0, 250.1, 250.2, 250.3	E10.0, E10.1, E10.9, E11.0, E11.1, E11.9, E12.0, E12.1, E12.9, E13.0, E13.1, E13.9, E14.0, E14.1, E14.9
Drug abuse	292, 304, 305.2, 305.3, 305.4, 305.5, 305.6, 305.7, 305.8, 305.9, V65.42	F11, F12, F13, F14, F15, F16, F18, F19, Z71.5, Z72.2
Fluid and electrolyte disorders	253.6, 276	E22.2, E86, E87
HIV/AIDS	042, 043, 044	B20, B21, B22, B24
Hypertension, with complications	402, 403, 404, 405	I11, I12, I13, I15
Hypertension, without complications	401	I10
Hypothyroidism	240.9, 243, 244, 246.1, 246.8	E00, E01, E02, E03, E89.0
Liver disease	070.22, 070.23, 070.32, 070.33, 070.44, 070.54, 070.6, 070.9, 456.0, 456.1, 456.2, 570, 571, 572.2, 572.3, 572.4, 572.8, 573.3, 573.4, 573.8, 573.9, V42.7	B18, I85, I86.4, I98.2, K70, K71.1, K71.3, K71.4, K71.5, K71.7, K72, K73, K74, K76.0, K76.2, K76.3, K76.4, K76.5, K76.6, K76.7, K76.8, K76.9, Z94.4
Lymphoma	200, 201, 202, 203.0, 238.6	C81, C82, C83, C84, C85, C88, C96, C90.0, C90.2
Metastatic cancer	196, 197, 198, 199	C77, C78, C79, C80
Neurological disorders, other	331.9, 332.0, 332.1, 333.4, 333.5, 333.92, 334, 335, 336.2, 340, 341, 345, 348.1, 348.3, 780.3, 784.3	G10, G11, G12, G13, G20, G21, G22, G25.4, G25.5, G31.2, G31.8, G31.9, G32, G35, G36, G37, G40, G41, G93.1, G93.4, R47.0, R56
Obesity	278.0	E66
Paralysis	334.1, 342, 343, 344.0, 344.1, 344.2, 344.3, 344.4, 344.5, 344.6, 344.9	G04.1, G11.4, G80.1, G80.2, G81, G82, G83.0, G83.1, G83.2, G83.3, G83.4, G83.9
Peptic ulcer disease (excluding bleeding)	531.7, 531.9, 532.7, 532.9, 533.7, 533.9, 534.7, 534.9	K25.7, K25.9, K26.7, K26.9, K27.7, K27.9, K28.7, K28.9
Peripheral vascular disorders	093.0, 437.3, 440, 441, 443.1, 443.2, 443.8, 443.9, 447.1, 557.1, 557.9, V43.4	I70, I71, I73.1, I73.8, I73.9, I77.1, I79.0, I79.2, K55.1, K55.8, K55.9, Z95.8, Z95.9
Psychoses	293.8, 295, 296.04, 296.14, 296.44, 296.54, 297, 298	F20, F22, F23, F24, F25, F28, F29, F30.2, F31.2, F31.5
Pulmonary circulation disorders	415.0, 415.1, 416, 417.0, 417.8, 417.9	I26, I27, I28.0, I28.8, I28.9
Renal failure	403.01, 403.11, 403.91, 404.02, 404.03, 404.12, 404.13, 404.92, 404.93, 585, 586, 588.0, V42.0, V45.1, V56	I12.0, I13.1, N18, N19, N25.0, Z49.0, Z49.1, Z49.2, Z94.0, Z99.2

Rheumatoid arthritis	446, 701.0, 710.0, 710.1, 710.2, 710.3, 710.4, 710.8, 710.9, 711.2, 714, 719.3, 720, 725, 728.5, 728.89, 729.30	L94.0, L94.1, L94.3, M05, M06, M08, M12.0, M12.3, M30, M31.0, M31.1, M31.2, M31.3, M32, M33, M34, M35, M45, M46.1, M46.8, M46.9
Solid tumor, without metastasis	140, 141, 142, 143, 144, 145, 146, 147, 148, 149, 150, 151, 152, 153, 154, 155, 156, 157, 158, 159, 160, 161, 162, 163, 164, 165, 166, 167, 168, 169, 170, 171, 172, 174, 175, 176, 177, 178, 179, 180, 181, 182, 183, 184, 185, 186, 187, 188, 189, 190, 191, 192, 193, 194, 195	C00, C01, C02, C03, C04, C05, C06, C07, C08, C09, C10, C11, C12, C13, C14, C15, C16, C17, C18, C19, C20, C21, C22, C23, C24, C25, C26, C30, C31, C32, C33, C34, C37, C38, C39, C40, C41, C43, C45, C46, C47, C48, C49, C50, C51, C52, C53, C54, C55, C56, C57, C58, C60, C61, C62, C63, C64, C65, C66, C67, C68, C69, C70, C71, C72, C73, C74, C75, C76, C97
Valvular disease	093.2, 394, 395, 396, 397, 424, 746.3, 746.4, 746.5, 746.6, V42.2, V43.3	A52.0, I05, I06, I07, I08, I09.1, I09.8, I34, I35, I36, I37, I38, I39, Q23.0, Q23.1, Q23.2, Q23.3, Z95.2, Z95.3, Z95.4
Weight loss	260, 261, 262, 263, 783.2, 799.4	E40, E41, E42, E43, E44, E45, E46, R63.4, R64

ICD = International Statistical Classification of Diseases and Related Health Problems, HIV/AIDS = human immunodeficiency virus/acquired immunodeficiency syndrome.

Appendix B. Frequency and prevalence of disease categories ascertained using the Elixhauser comorbidity index.

Chronic disease category	Frequency	Prevalence (%)
Hypertension, without complications	338,647	22.42
Chronic pulmonary disease	231,748	15.34
Depression	191,666	12.69
Diabetes, without complications	146,939	9.73
Deficiency anemia	94,175	6.23
Hypothyroidism	90,843	6.01
Obesity	70,856	4.69
Cardiac arrhythmia	67,772	4.49
Solid tumor, without metastasis	65,667	4.35
Neurological disorders, other	41,243	2.73
Congestive heart failure	40,346	2.67
Diabetes, with complications	39,268	2.60
Fluid and electrolyte disorders	37,678	2.49
Liver disease	34,125	2.26
Psychoses	33,303	2.20
Rheumatoid arthritis	32,776	2.17
Peripheral vascular disorders	27,797	1.84
Renal failure	23,130	1.53
Alcohol abuse	21,798	1.44
Weight loss	18,315	1.21
Drug abuse	17,490	1.16
Coagulopathy	16,154	1.07
Metastatic cancer	13,606	0.90
Valvular disease	13,023	0.86
Peptic ulcer disease (excluding bleeding)	10,236	0.68
Pulmonary circulation disorders	7,660	0.51
Lymphoma	6,991	0.46
Paralysis	6,502	0.43
Hypertension, with complications	4,511	0.30
Blood loss anemia	3,767	0.25
HIV/AIDS	1,771	0.12

HIV/AIDS = human immunodeficiency virus/acquired immunodeficiency syndrome.

Appendix C. Frequency and prevalence of chronic disease categories ascertained using the Johns Hopkins Adjusted Clinical Groups (ACG) System.

High-level disease category	Chronic disease category	Frequency	Prevalence (%)	Excluded
Administrative	Administrative concerns and non-specific laboratory abnormalities	423	0.03	✓
	Surgical aftercare	366	0.02	✓
	Transplant status	1,221	0.08	✓
Allergy	Asthma	136,609	9.04	
	Disorders of the immune system	5,743	0.38	
Cardiovascular	Acute myocardial infarction	1,315	0.09	
	Cardiac arrest, shock	943	0.06	
	Cardiac arrhythmia	61,786	4.09	
	Cardiac valve disorders	11,836	0.78	
	Cardiomyopathy	6,769	0.45	
	Cardiovascular disorders, other	17,018	1.13	
	Cardiovascular signs and symptoms	2,639	0.17	
	Congenital heart disease	1,989	0.13	
	Congestive heart failure	38,157	2.53	
	Disorders of lipid metabolism	147,942	9.79	
	Generalized atherosclerosis	22,978	1.52	
	Heart murmur	<6	<0.01	✓
	Hypertension	340,572	22.54	
	Ischemic heart disease (excluding acute myocardial infarction)	72,660	4.81	
Dental	Disorders of mouth	680	0.05	
Ear, Nose, Throat	Chronic pharyngitis and tonsillitis	16,672	1.1	
	Deafness, hearing loss	33,015	2.19	
	ENT disorders, other	2,570	0.17	
	Otitis externa	168	0.01	
	Otitis media	464	0.03	
	Temporomandibular joint disease	9,402	0.62	
Endocrine	Hypothyroidism	89,286	5.91	
	Osteoporosis	21,384	1.42	
	Other endocrine disorders	43,315	2.87	
	Short stature	798	0.05	✓
	Type 1 diabetes	14,237	0.94	
	Type 2 diabetes	136,611	9.04	
Eye	Age-related macular degeneration	12,870	0.85	
	Blindness	3,929	0.26	
	Cataract, aphakia	66,427	4.4	
	Conjunctivitis, keratitis	4,199	0.28	
	Diabetic retinopathy	12,537	0.83	
	Disorders of the eyelid and lacrimal duct	8,636	0.57	
	Eye, other disorders	20,487	1.36	
	Glaucoma	42,365	2.8	
	Ophthalmic signs and symptoms	1,203	0.08	
	Refractive errors	539,941	35.74	✓
	Retinal disorders (excluding diabetic retinopathy)	11,055	0.73	

	Strabismus, amblyopia	4,367	0.29	✓
Female Reproductive	Endometriosis	9,787	0.65	
	Female gynecologic conditions, other	12	<0.01	✓
Gastrointestinal/Hepatic	Acute hepatitis	118	<0.01	✓
	Chronic liver disease	23,151	1.53	
	Chronic pancreatitis	3,967	0.26	
	Diverticular disease of colon	35,755	2.37	
	Gastroenteritis	36	<0.01	
	Gastroesophageal reflux	5,568	0.37	
	Gastrointestinal signs and symptoms	2,311	0.15	
	Gastrointestinal/hepatic disorders, other	13,607	0.9	
	Hepatitis C	1,386	0.09	
	Inflammatory bowel disease	13,975	0.93	
	Irritable bowel syndrome	28,283	1.87	
	Lactose intolerance	1,529	0.1	
General Signs and Symptoms	Debility and undue fatigue	8,503	0.56	
	Lymphadenopathy	<6	<0.01	✓
	Nonspecific signs and symptoms	1,231	0.08	✓
General Surgery	Anorectal conditions	1,540	0.1	
	Aortic aneurysm	1,552	0.1	
	Benign and unspecified neoplasm	36,277	2.4	
	Cholelithiasis, cholecystitis	21,440	1.42	
	Chronic cystic disease of the breast	3,669	0.24	
	Gastrointestinal obstruction/perforation	46	<0.01	
	Peripheral vascular disease	20,448	1.35	
	Varicose veins of lower extremities	18,603	1.23	
Genetic	Chromosomal anomalies	2,619	0.17	
	Inherited metabolic disorders	44,445	2.94	
Genito-urinary	Genito-urinary disorders, other	7,713	0.51	
	Incontinence	11,002	0.73	
	Other male genital disease	614	0.04	
	Prostatic hypertrophy	32,176	2.13	
	Prostatitis	1,183	0.08	
	Renal calculi	749	0.05	
	Urinary symptoms	635	0.04	
	Urinary tract infections	1,510	0.1	
	Vesicoureteral reflux	2,189	0.14	
Hematologic	Aplastic anemia	1,302	0.09	
	Deep vein thrombosis	369	0.02	
	Deficiency anemias	42,885	2.84	
	Hematologic disorders, other	3,688	0.24	
	Hemophilia, coagulation disorder	6,984	0.46	
	Neonatal jaundice	111	<0.01	✓
	Other hemolytic anemias	2,996	0.2	
	Sickle cell disease	531	0.04	
	Thrombophlebitis	14	<0.01	✓
Infections	Fungal infections	26	<0.01	

	HIV, AIDS	1,816	0.12	
	Infections, other	271	0.02	
	Sexually transmitted diseases	300	0.02	
	Tuberculosis infection	<6	<0.01	✓
Malignancies	Acute leukemia	1,233	0.08	
	High impact malignant neoplasms	19,059	1.26	✓
	Low impact malignant neoplasms	19,812	1.31	✓
	Malignant neoplasms of the skin	12,626	0.84	
	Malignant neoplasms, bladder	3,883	0.26	
	Malignant neoplasms, breast	11,789	0.78	
	Malignant neoplasms, cervix, uterus	3,549	0.23	
	Malignant neoplasms, colorectal	9,859	0.65	
	Malignant neoplasms, esophagus	924	0.06	
	Malignant neoplasms, kidney	2,882	0.19	
	Malignant neoplasms, liver and biliary tract	1,816	0.12	
	Malignant neoplasms, lung	7,724	0.51	
	Malignant neoplasms, lymphomas	6,077	0.4	
	Malignant neoplasms, ovary	1,791	0.12	
	Malignant neoplasms, pancreas	1,713	0.11	
	Malignant neoplasms, prostate	11,072	0.73	
	Malignant neoplasms, stomach	1,328	0.09	
Musculoskeletal	Acquired foot deformities	3,307	0.22	
	Acute sprains and strains	848	0.06	✓
	Amputation status	256	0.02	✓
	Bursitis, synovitis, tenosynovitis	8,254	0.55	
	Cervical pain syndromes	5,639	0.37	
	Congenital anomalies of limbs, hands, and feet	93	<0.01	✓
	Congenital hip dislocation	16	<0.01	✓
	Degenerative joint disease	137,076	9.07	
	Fracture of neck of femur (hip)	16	<0.01	✓
	Fractures (excluding digits)	34	<0.01	✓
	Joint disorders, trauma related	14,288	0.95	
	Kyphoscoliosis	2,535	0.17	
	Low back pain	34,195	2.26	
	Musculoskeletal disorders, other	21,863	1.45	
	Musculoskeletal signs and symptoms	1,975	0.13	
Neonatal	Disorders of newborn period	205	0.01	✓
	Newborn status, complicated	70	<0.01	✓
Neurologic	Autism Spectrum Disorder	5,122	0.34	
	Central nervous system infections	411	0.03	
	Cerebral palsy	1,971	0.13	
	Cerebrovascular disease	31,280	2.07	
	Delirium	252	0.02	
	Dementia	29,036	1.92	
	Developmental disorder	11,459	0.76	
	Head injury	178	0.01	
	Migraines	1,419	0.09	

	Multiple sclerosis	4,480	0.3	
	Muscular dystrophy	2,231	0.15	
	Neurologic disorders, other	20,979	1.39	
	Neurologic signs and symptoms	5,252	0.35	
	Organic brain syndrome	12,393	0.82	
	Paralytic syndromes, other	2,916	0.19	
	Parkinsons disease	5,746	0.38	
	Peripheral neuropathy, neuritis	15,342	1.02	
	Quadriplegia and paraplegia	1,000	0.07	
	Seizure disorder	12,642	0.84	
	Sleep problems	719	0.05	
	Spinal cord injury/disorders	10,474	0.69	
	Vertiginous syndromes	99	<0.01	
Nutrition	Failure to thrive	15,003	0.99	
	Nutritional deficiencies	552	0.04	
	Nutritional disorders, other	3,256	0.22	
	Obesity	68,231	4.52	
Psychosocial	Anxiety, neuroses	73,751	4.88	
	Attention deficit disorder	30,692	2.03	
	Bipolar disorder	11,770	0.78	
	Depression	167,133	11.06	
	Eating disorder	809	0.05	
	Impulse control	256	0.02	
	Major depression	28,737	1.9	
	Personality disorders	8,957	0.59	
	Post traumatic stress disorder	6,465	0.43	
	Psychologic signs and symptoms	1,181	0.08	
	Psychological disorders of childhood	6,303	0.42	
	Psychosexual	1,797	0.12	
	Psych-physiologic and somatoform disorders	3,578	0.24	
	Schizophrenia and affective psychosis	10,633	0.7	
	Sleep disorders of nonorganic origin	594	0.04	
	Substance use	15,414	1.02	
Reconstructive	Chronic ulcer of the skin	16,369	1.08	
	Cleft lip and palate	197	0.01	
Renal	Accute renal failure	21	<0.01	✓
	Chronic renal failure	15,592	1.03	
	ESRD	3,250	0.22	
	Fluid/electrolyte disturbances	834	0.06	
	Nephritis, nephrosis	3,345	0.22	
	Renal disorders, other	22,647	1.5	
Respiratory	Acute lower respiratory tract infection	12	<0.01	✓
	Acute respiratory failure	259	0.02	✓
	Chronic respiratory failure	8,969	0.6	
	Cystic fibrosis	399	0.03	
	Emphysema, chronic bronchitis, COPD	55,611	3.68	
	Pulmonary embolism	590	0.04	

	Respiratory disorders, other	33,318	2.21	
	Respiratory signs and symptoms	1,271	0.08	
	Sinusitis	12	<0.01	✓
	Sleep apnea	40,007	2.65	
	Tracheostomy	220	0.01	✓
Rheumatologic	Arthropathy	5,014	0.33	
	Autoimmune and connective tissue diseases	20,356	1.35	
	Gout	33,846	2.24	
	Raynauds syndrome	85	<0.01	
	Rheumatoid arthritis	16,132	1.07	
Skin	Other inflammatory conditions of skin	11	<0.01	✓
	Other skin disorders	1,025	0.07	
	Psoriasis	19,452	1.29	
Toxic Effects and Adverse Events	Adverse effects of medicinal agents	194	0.01	✓
	Adverse events from medical/surgical procedures	397	0.03	✓
	Complications of mechanical devices	230	0.02	✓
	Toxic effects of nonmedicinal agents	33	<0.01	✓

ENT = ear, nose, and throat; HIV/AIDS = human immunodeficiency virus/acquired immunodeficiency syndrome; ESRD = end-stage renal disease; COPD = chronic obstructive pulmonary disease.

Appendix D. Frequency and prevalence of Expanded Diagnostic Clusters, ascertained using the Johns Hopkins Adjusted Clinical Groups (ACG) System, stratified by sex.

Expanded Diagnostic Cluster	Male (N=756,198)		Female (N=754,480)	
	N	%	N	%
Acquired foot deformities	1,087	0.14	2,220	0.29
Acute leukemia	703	0.09	530	0.07
Acute myocardial infarction	782	0.1	533	0.07
Acute respiratory failure	135	0.02	124	0.02
Age-related macular degeneration	4,971	0.66	7,899	1.05
Anorectal conditions	760	0.1	780	0.1
Anxiety, neuroses	26,488	3.5	47,261	6.26
Aortic aneurysm	954	0.13	598	0.08
Aplastic anemia	699	0.09	603	0.08
Arthropathy	2,063	0.27	2,951	0.39
Asthma	62,134	8.22	74,474	9.87
Attention deficit disorder	20,454	2.7	10,237	1.36
Autism spectrum disorder	3,970	0.52	1,152	0.15
Autoimmune and connective tissue diseases	6,932	0.92	13,424	1.78
Benign and unspecified neoplasm	15,469	2.05	20,808	2.76
Bipolar disorder	4,772	0.63	6,998	0.93
Blindness	1,811	0.24	2,118	0.28
Bursitis, synovitis, tenosynovitis	3,571	0.47	4,683	0.62
Cardiac arrest, shock	544	0.07	399	0.05
Cardiac arrhythmia	33,174	4.39	28,612	3.79
Cardiac valve disorders	6,415	0.85	5,421	0.72
Cardiomyopathy	4,359	0.58	2,410	0.32
Cardiovascular disorders, other	7,502	0.99	9,516	1.26
Cardiovascular signs and symptoms	1,393	0.18	1,246	0.17
Cataract, aphakia	28,320	3.75	38,107	5.05
Central nervous system infections	217	0.03	194	0.03
Cerebral palsy	1,063	0.14	908	0.12
Cerebrovascular disease	15,378	2.03	15,901	2.11
Cervical pain syndromes	2,796	0.37	2,842	0.38
Cholelithiasis, cholecystitis	6,856	0.91	14,584	1.93
Chromosomal anomalies	1,073	0.14	1,544	0.2
Chronic cystic disease of the breast	36	<0.01	3,633	0.48
Chronic liver disease	12,171	1.61	10,979	1.46
Chronic pancreatitis	1,786	0.24	2,181	0.29
Chronic pharyngitis and tonsillitis	7,280	0.96	9,392	1.24
Chronic renal failure	8,369	1.11	7,223	0.96
Chronic ulcer of the skin	8,430	1.11	7,939	1.05
Cleft lip and palate	102	0.01	95	0.01
Congenital heart disease	1,053	0.14	934	0.12
Congestive heart failure	19,007	2.51	19,150	2.54
Conjunctivitis, keratitis	1,625	0.21	2,574	0.34
Cystic fibrosis	203	0.03	196	0.03

Deafness, hearing loss	16,485	2.18	16,529	2.19
Debility and undue fatigue	3,087	0.41	5,416	0.72
Deep vein thrombosis	180	0.02	189	0.03
Deficiency anemias	15,780	2.09	27,105	3.59
Degenerative joint disease	56,654	7.49	80,422	10.66
Delirium	116	0.02	136	0.02
Dementia	11,560	1.53	17,475	2.32
Depression	58,476	7.73	108,651	14.4
Developmental disorder	7,141	0.94	4,317	0.57
Diabetic retinopathy	6,484	0.86	6,053	0.8
Disorders of lipid metabolism	79,433	10.5	68,507	9.08
Disorders of mouth	259	0.03	421	0.06
Disorders of the eyelid and lacrimal duct	3,033	0.4	5,603	0.74
Disorders of the immune system	2,655	0.35	3,088	0.41
Diverticular disease of colon	16,697	2.21	19,058	2.53
Eating disorder	104	0.01	704	0.09
Emphysema, chronic bronchitis, COPD	27,029	3.57	28,582	3.79
Endometriosis	0	0	9,779	1.3
ENT disorders, other	1,104	0.15	1,466	0.19
ESRD	1,860	0.25	1,390	0.18
Eye, other disorders	9,556	1.26	10,931	1.45
Failure to thrive	7,538	1	7,464	0.99
Female gynecologic conditions, other	0	0	12	<0.01
Fluid/electrolyte disturbances	333	0.04	501	0.07
Fungal infections	10	<0.01	16	<0.01
Gastroenteritis	16	<0.01	20	<0.01
Gastroesophageal reflux	3,279	0.43	2,289	0.3
Gastrointestinal obstruction/perforation	27	<0.01	19	<0.01
Gastrointestinal signs and symptoms	990	0.13	1,320	0.17
Gastrointestinal/hepatic disorders, other	5,344	0.71	8,262	1.1
Generalized atherosclerosis	12,369	1.64	10,608	1.41
Genito-urinary disorders, other	2,222	0.29	5,486	0.73
Glaucoma	17,479	2.31	24,886	3.3
Gout	24,187	3.2	9,659	1.28
Head injury	109	0.01	69	<0.01
Heart murmur	0	0	<6	<0.01
Hematologic disorders, other	1,565	0.21	2,123	0.28
Hemophilia, coagulation disorder	3,307	0.44	3,677	0.49
Hepatitis C	805	0.11	581	0.08
HIV, AIDS	1,149	0.15	667	0.09
Hypertension	167,982	22.21	172,588	22.88
Hypothyroidism	21,834	2.89	67,452	8.94
Impulse control	136	0.02	120	0.02
Incontinence	<6	<0.01	11,000	1.46
Infections, other	119	0.02	152	0.02
Inflammatory bowel disease	6,217	0.82	7,758	1.03
Inherited metabolic disorders	23,309	3.08	21,136	2.8

Irritable bowel syndrome	8,655	1.14	19,628	2.6
Ischemic heart disease (excluding acute myocardial infarction)	44,168	5.84	28,492	3.78
Joint disorders, trauma related	7,178	0.95	7,110	0.94
Kyphoscoliosis	807	0.11	1,728	0.23
Lactose intolerance	690	0.09	839	0.11
Low back pain	16,384	2.17	17,810	2.36
Lymphadenopathy	<6	<0.01	<6	<0.01
Major depression	10,299	1.36	18,436	2.44
Malignant neoplasms of the skin	6,773	0.9	5,853	0.78
Malignant neoplasms, bladder	2,794	0.37	1,089	0.14
Malignant neoplasms, breast	154	0.02	11,635	1.54
Malignant neoplasms, cervix, uterus	0	0	3,549	0.47
Malignant neoplasms, colorectal	5,218	0.69	4,641	0.62
Malignant neoplasms, esophagus	640	0.08	284	0.04
Malignant neoplasms, kidney	1,837	0.24	1,045	0.14
Malignant neoplasms, liver and biliary tract	1,020	0.13	796	0.11
Malignant neoplasms, lung	3,749	0.5	3,975	0.53
Malignant neoplasms, lymphomas	3,303	0.44	2,774	0.37
Malignant neoplasms, ovary	0	0	1,791	0.24
Malignant neoplasms, pancreas	839	0.11	874	0.12
Malignant neoplasms, prostate	11,072	1.46	0	0
Malignant neoplasms, stomach	784	0.1	544	0.07
Migraines	317	0.04	1,102	0.15
Multiple sclerosis	1,386	0.18	3,094	0.41
Muscular dystrophy	1,210	0.16	1,021	0.14
Musculoskeletal disorders, other	10,839	1.43	11,024	1.46
Musculoskeletal signs and symptoms	891	0.12	1,084	0.14
Nephritis, nephrosis	1,712	0.23	1,633	0.22
Neurologic disorders, other	9,299	1.23	11,680	1.55
Neurologic signs and symptoms	2,367	0.31	2,885	0.38
Nutritional deficiencies	232	0.03	320	0.04
Nutritional disorders, other	1,579	0.21	1,677	0.22
Obesity	28,072	3.71	40,158	5.32
Ophthalmic signs and symptoms	534	0.07	669	0.09
Organic brain syndrome	4,967	0.66	7,425	0.98
Osteoporosis	2,629	0.35	18,755	2.49
Other endocrine disorders	12,640	1.67	30,675	4.07
Other hemolytic anemias	1,346	0.18	1,650	0.22
Other inflammatory conditions of skin	<6	<0.01	7	<0.01
Other male genital disease	614	0.08	0	0
Other skin disorders	366	0.05	659	0.09
Otitis externa	69	<0.01	99	0.01
Otitis media	205	0.03	259	0.03
Paralytic syndromes, other	1,540	0.2	1,375	0.18
Parkinsons disease	3,160	0.42	2,586	0.34
Peripheral neuropathy, neuritis	7,199	0.95	8,143	1.08
Peripheral vascular disease	12,062	1.6	8,386	1.11

Personality disorders	3,267	0.43	5,688	0.75
Post traumatic stress disorder	2,429	0.32	4,036	0.53
Prostatic hypertrophy	32,176	4.25	0	0
Prostatitis	1,181	0.16	0	0
Psoriasis	9,424	1.25	10,028	1.33
Psychologic signs and symptoms	835	0.11	346	0.05
Psychological disorders of childhood	3,770	0.5	2,533	0.34
Psychosexual	1,039	0.14	748	0.1
Psych-physiologic and somatoform disorders	1,334	0.18	2,244	0.3
Pulmonary embolism	269	0.04	321	0.04
Quadriplegia and paraplegia	665	0.09	335	0.04
Raynauds syndrome	21	<0.01	64	<0.01
Renal calculi	435	0.06	314	0.04
Renal disorders, other	12,244	1.62	10,403	1.38
Respiratory disorders, other	15,623	2.07	17,695	2.35
Respiratory signs and symptoms	647	0.09	624	0.08
Retinal disorders (excluding diabetic retinopathy)	5,217	0.69	5,838	0.77
Rheumatoid arthritis	4,678	0.62	11,454	1.52
Schizophrenia and affective psychosis	6,113	0.81	4,520	0.6
Seizure disorder	6,460	0.85	6,182	0.82
Sexually transmitted diseases	134	0.02	166	0.02
Sickle cell disease	248	0.03	283	0.04
Sinusitis	8	<0.01	<6	<0.01
Sleep apnea	24,305	3.21	15,701	2.08
Sleep disorders of nonorganic origin	218	0.03	375	0.05
Sleep problems	403	0.05	316	0.04
Spinal cord injury/disorders	5,505	0.73	4,967	0.66
Substance use	8,654	1.14	6,760	0.9
Temporomandibular joint disease	3,107	0.41	6,293	0.83
Thrombophlebitis	10	<0.01	<6	<0.01
Tuberculosis infection	<6	<0.01	<6	<0.01
Type 1 diabetes	7,502	0.99	6,735	0.89
Type 2 diabetes	71,138	9.41	65,473	8.68
Urinary symptoms	318	0.04	317	0.04
Urinary tract infections	261	0.03	1,249	0.17
Varicose veins of lower extremities	5,368	0.71	13,235	1.75
Vertiginous syndromes	19	<0.01	80	0.01
Vesicoureteral reflux	1,267	0.17	922	0.12

ENT = ear, nose, and throat; HIV/AIDS = human immunodeficiency virus/acquired immunodeficiency syndrome; ESRD = end-stage renal disease; COPD = chronic obstructive pulmonary disease.

Appendix E. Demographic and healthcare utilization characteristics of Manitoba residents (2015/16-2018/19) stratified by number of chronic conditions (*N*=1,510,678).

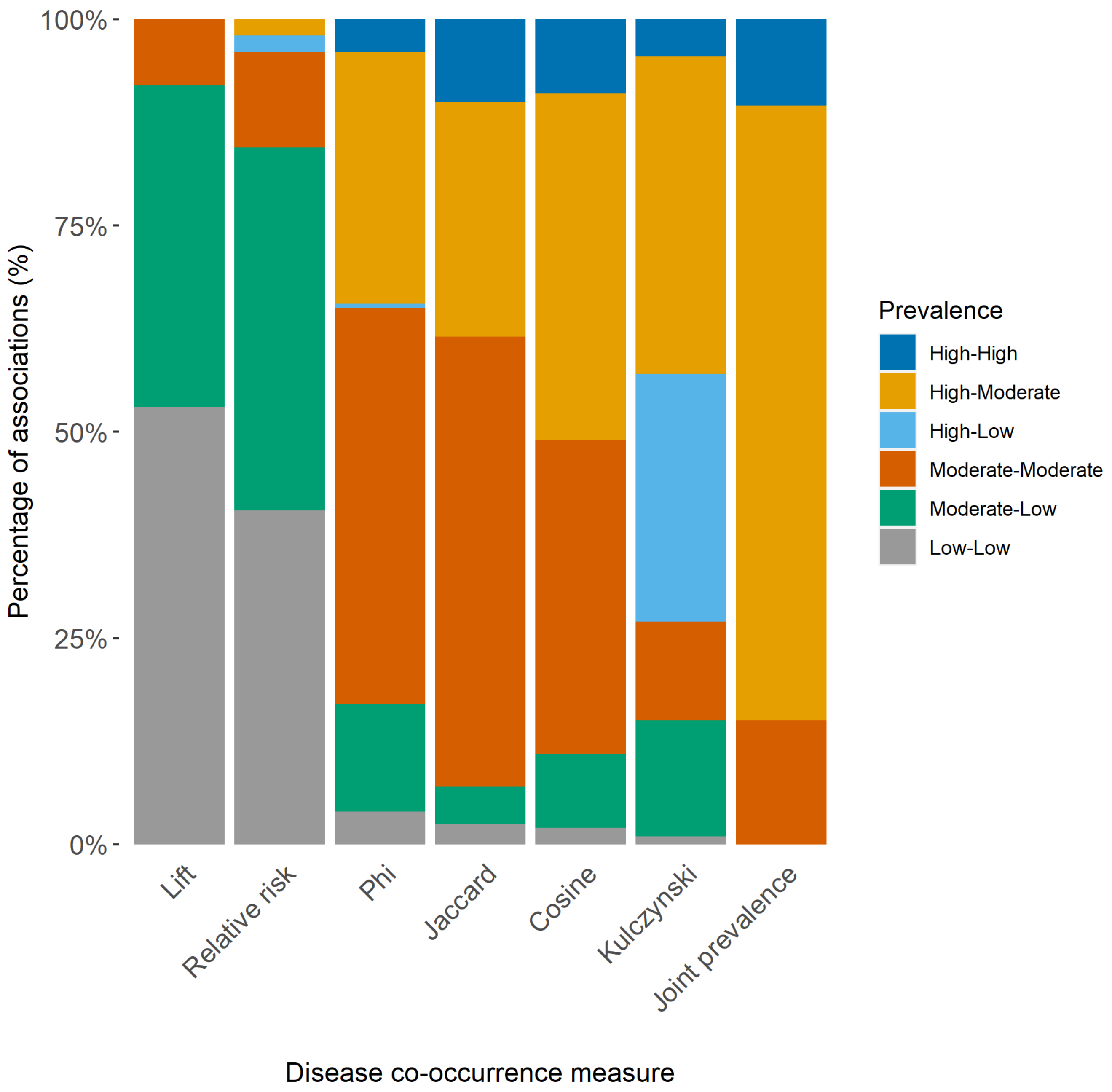
	Number of chronic conditions	
	<2 *n*=892,767 (59.1%)	2+ *n*=617,911 (40.9%)
Sex		
Male	472,524 (52.9)	283,674 (45.9)
Female	420,243 (47.1)	334,237 (54.1)
Age (years)		
<20	323,359 (36.2)	43,072 (7.0)
20-39	326,582 (36.6)	102,750 (16.6)
49-59	182,116 (20.4)	189,300 (30.6)
60+	60,710 (6.8)	282,789 (45.8)
Residence locality		
Rural	345,109 (38.7)	221,923 (35.9)
Urban	547,080 (61.3)	395,907 (64.1)
Unknown	578 (0.1)	81 (<0.1)
Income quintile		
Q1 (lowest)	188,982 (21.2)	120,654 (19.5)
Q2	175,021 (19.6)	121,899 (19.7)
Q3	170,872 (19.1)	127,697 (20.7)
Q4	177,267 (19.9)	119,901 (19.4)
Q5 (highest)	175,741 (19.7)	115,384 (18.7)
Unknown	4,884 (0.6)	12,376 (2.0)
Healthcare utilization		
Inpatient hospitalization	36,619 (4.1)	83,934 (13.6)
Ambulatory visits	1 (0-3)	6 (3-10)

Data are presented as N (%) or median (Q1-Q3).

Demographic characteristics were measured at exit date.

Healthcare utilization was measured during the last 12 months of follow-up.

Appendix F. Percentage of the strongest 200 statistically significant pairwise chronic disease co-occurrence relationships characterized by prevalence, among select co-occurrence measures.



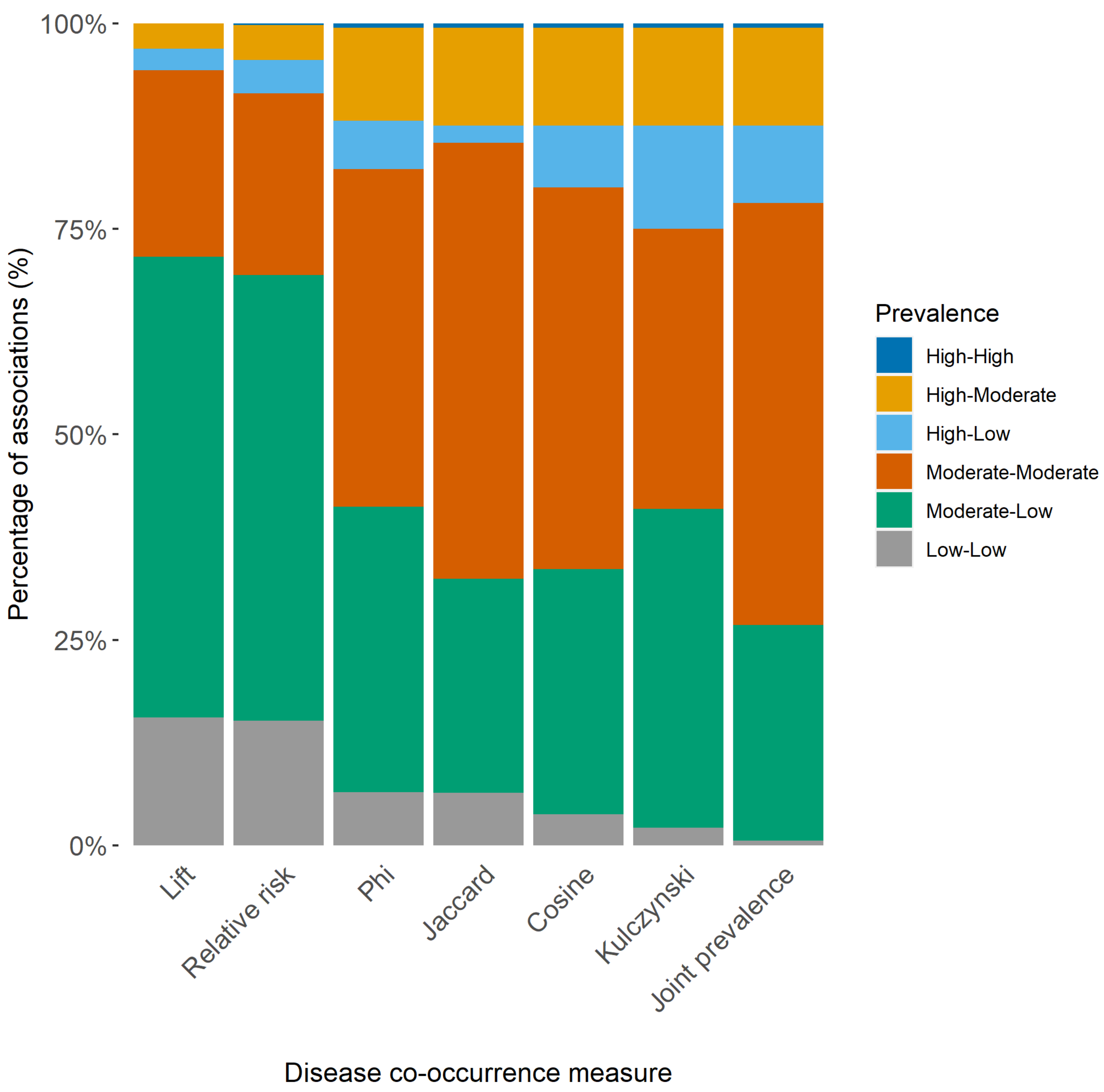
Note: Chronic diseases were ascertained using the Johns Hopkins ACG System; prevalence was categorized as low (<0.5%), moderate (0.5 to <5%), and high (≥5%).

Appendix G. Number (%) of the strongest 200 statistically significant pairwise chronic disease co-occurrence relationships characterized by prevalence, among select co-occurrence measures.

Prevalence	Lift	Relative risk	Phi	Jaccard	Cosine	Kulczynski	Joint prevalence
High-High	0 (0.0)	0 (0.0)	8 (4.0)	20 (10.0)	18 (9.0)	9 (4.5)	21 (10.5)
High-Moderate	0 (0.0)	4 (2.0)	61 (30.5)	57 (28.5)	84 (42.0)	77 (38.5)	149 (74.5)
High-Low	0 (0.0)	4 (2.0)	1 (0.5)	0 (0.0)	0 (0.0)	60 (30.0)	0 (0.0)
Moderate-Moderate	16 (8.0)	23 (11.5)	96 (48.0)	109 (54.5)	76 (38.0)	24 (12.0)	30 (15.0)
Moderate-Low	78 (39.0)	88 (44.0)	26 (13.0)	9 (4.5)	18 (9.0)	28 (14.0)	0 (0.0)
Low-Low	106 (53.0)	81 (40.5)	8 (4.0)	5 (2.5)	4 (2.0)	2 (1.0)	0 (0.0)

Note: chronic diseases were ascertained using the Johns Hopkins ACG System; prevalence was categorized as low (<0.5%), moderate (0.5 to <5%), and high (≥5%).

Appendix H. Percentage of the strongest 50 percent (*n*=3,922) of statistically significant pairwise chronic disease co-occurrence relationships characterized by prevalence, among select co-occurrence measures.



Note: Chronic diseases were ascertained using the Johns Hopkins ACG System; and prevalence was categorized as low (<0.5%), moderate (0.5 to <5%), and high (≥5%).

Appendix I. Number (%) of the strongest 50 percent (*n*=3,922) of statistically significant pairwise chronic disease co-occurrence relationships characterized by prevalence, among select co-occurrence measures.

Prevalence	Lift	Relative risk	Phi	Jaccard	Cosine	Kulczynski	Joint prevalence
High-High	1 (0.0)	8 (0.2)	21 (0.5)	21 (0.5)	21 (0.5)	21 (0.5)	21 (0.5)
High-Moderate	119 (3.0)	166 (4.2)	444 (11.3)	468 (11.9)	468 (11.9)	468 (11.9)	468 (11.9)
High-Low	104 (2.7)	161 (4.1)	231 (5.9)	81 (2.1)	294 (7.5)	492 (12.5)	367 (9.4)
Moderate-Moderate	889 (22.7)	865 (22.1)	1,610 (41.1)	2,079 (53.0)	1,822 (46.5)	1,336 (34.1)	2,016 (51.4)
Moderate-Low	2,198 (56.0)	2,128 (54.3)	1,362 (34.7)	1,021 (26.0)	1,168 (29.8)	1,522 (38.8)	1,027 (26.2)
Low-Low	611 (15.6)	594 (15.1)	254 (6.5)	252 (6.4)	149 (3.8)	83 (2.1)	23 (0.6)

Note: Chronic diseases were ascertained using the Johns Hopkins ACG System; prevalence was categorized as low (<0.5%), moderate (0.5 to <5%), and high (≥5%).