UNIVERSITY OF
EASTERN FINLAND

Clustering of medical diagnoses

Sami Sieranoja

19.12.2018

Introduction

Multimorbidity:

- Same patient has multiple diagnoses / diseases.

Example: Diabetes and heart disease.

- Majority of costs in health care come from multimorbid patients
- Clustering of diagnoses could help understand multimorbidity.
- Better coordination of treatments for different diseases.

Patient data

Patient	Visit	Disease code	Disease
1234	2009-10-5	G47	Sleep disorders
1234	2009-12-22	G35	Multiple sclerosis
1234	2009-4-28	I10	Essential (primary) hypertension
1234	2009-6-12	J45	Asthma
1234	2009-6-22	J45	Asthma
1234	2009-7-25	C73	Malignant neoplasm of thyroid gland
1234	2010-10-11	E11	Type 2 diabetes mellitus
1234	2011-6-10	F22	Delusional disorders
1234	2010-12-3	N30	Cystitis
1234	2010-4-13	S19	Other specified and unspecified injuries of neck
1234	2011-9-7	I10	Essential (primary) hypertension
1234	2012-1-7	E11	Type 2 diabetes mellitus
1234	2012-1-10	F10	Alcohol related disorders
1234	2013-10-15	K70	Alcoholic liver disease

N = 9149, Total number of patients

Key concepts

Search: sleep disorders

chisq	o/e	observed	expected	num(A)	num(B)	p(A)	p(B)	A	B
147.52	11.75	15	1.28	365	32	0.0399	0.0035	G47, Sleep disorders	J96, Respiratory failure, not elsewhere classified
524.34	5.12	158	30.84	773	365	0.0845	0.0399	E66, Overweight and obesity	G47, Sleep disorders
28.35	3.93	13	3.31	83	365	0.0091	0.0399	F31, Bipolar disorder	G47, Sleep disorders
257.16	3.58	138	38.5	365	965	0.0399	0.1055	G47, Sleep disorders	J45, Asthma
55.8	3.34	34	10.17	365	255	0.0399	0.0279	G47, Sleep disorders	J44, Other chronic obstructive pulmonary disease

N = 9149, Total number of patients
num(A) = Number of patients having diagnosis A

Example: A = Sleep disorders, B = Asthma
num(A) = 365 have Sleep disorders
num(B) = 965 have Asthma

Probability of Sleep disorders, p(A) = 365/9149 = 4%
Probability of asthma, p(B) = 965/9149 = 10%

Key concepts: Expected vs. Observed

Search: sleep disorders

chisq	o/e	observed	expected	num(A)	num(B)	p(A)	p(B)	A	B
147.52	11.75	15	1.28	365	32	0.0399	0.0035	G47, Sleep disorders	J96, Respiratory failure, not elsewhere classified
524.34	5.12	158	30.84	773	365	0.0845	0.0399	E66, Overweight and obesity	G47, Sleep disorders
28.35	3.93	13	3.31	83	365	0.0091	0.0399	F31, Bipolar disorder	G47, Sleep disorders
257.16	3.58	138	38.5	365	965	0.0399	0.1055	G47, Sleep disorders	J45, Asthma
55.8	3.34	34	10.17	365	255	0.0399	0.0279	G47, Sleep disorders	J44, Other chronic obstructive pulmonary disease

Assuming statistical independence, expected number of patients with both asthma and sleep disorders:

expected = $p(A) \cdot p(B) \cdot N$ = 10%*4%*9149 = 38

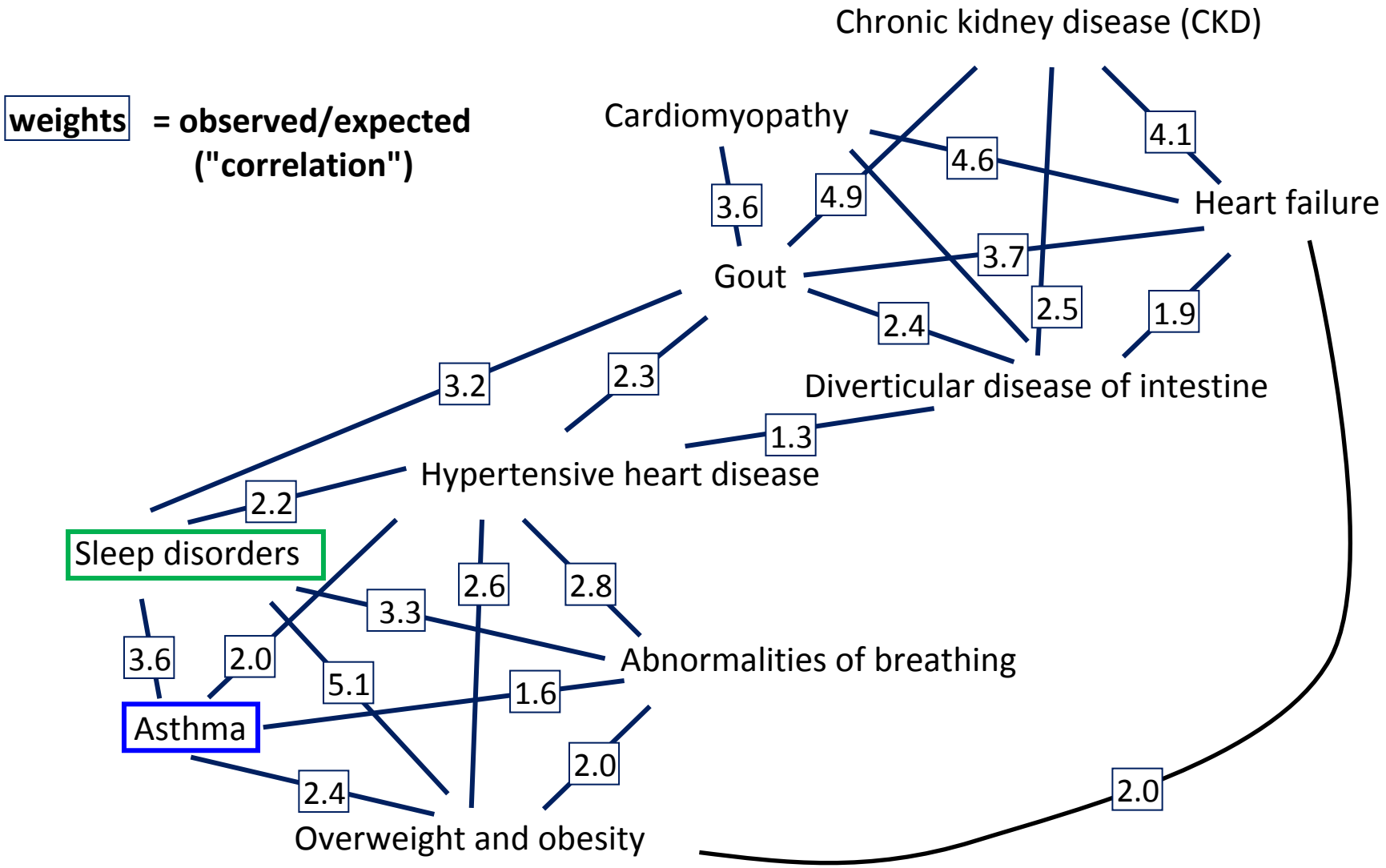
Actual observed number in data:

observed = 138

o/e=observed/expected = 138/38 = 3.6 times higher than expected

Graph of diagnoses

(Subgraph of the 237 diagnoses)



Objective function for clustering

W = sum of weights inside cluster

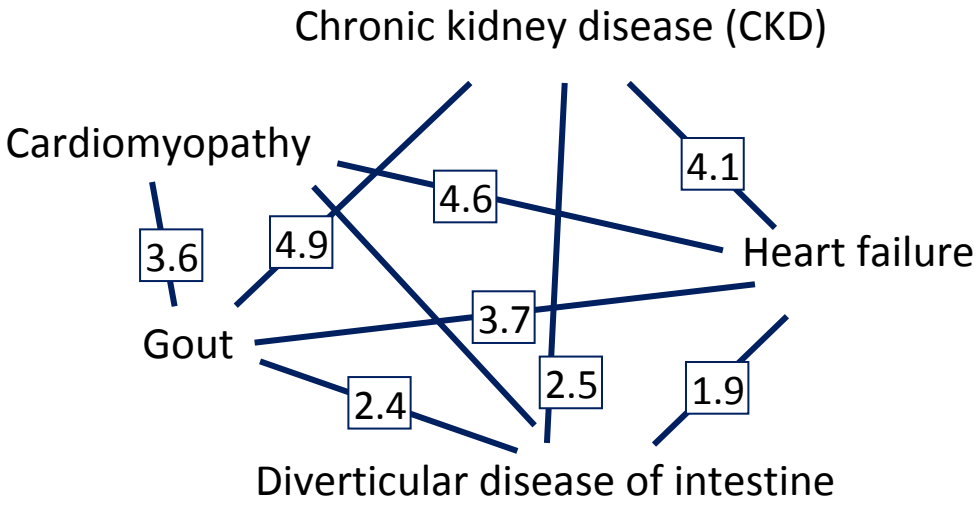
n = number of pairwise links

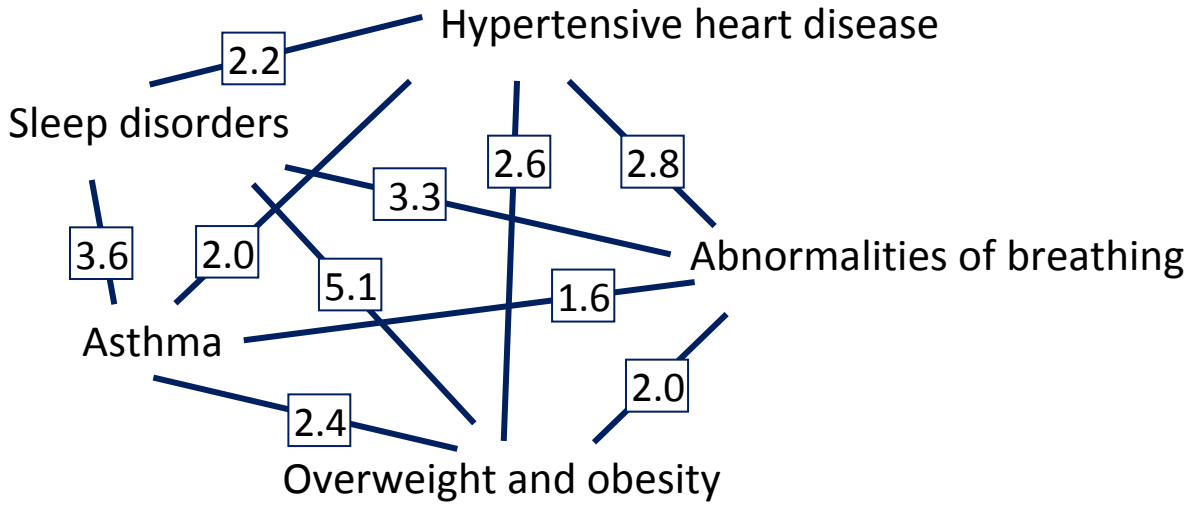
f = SUM(*W/n)* = Average weight inside all clusters

Maximize *f* using iterative optimization (work in progress)

Resulting clusters (2 out of 30)

weights **= observed/expected ("correlation")**





Result cluster 2/10

CLUSTER 2

	J96	G47	E66	J44	I11	J45	F17	E03	F10	F31	I87
J96		11.75	9.25	13.45	8.75	5.04					
G47	11.75		5.12	3.34	2.15	3.58		2.08		3.93	2.57
E66	9.25	5.12		2.69	2.58	2.37	2.93	1.88	1.56	2.14	2.14
J44	13.45	3.34	2.69		2.2	3.49	9.21	1.31	3.46		
I11	8.75	2.15	2.58	2.2		2.03	1.98	1.73			2.79
J45	5.04	3.58	2.37	3.49	2.03		1.59	1.8	1.25	1.71	1.79
F17			2.93	9.21	1.98	1.59			9.23		
E03		2.08	1.88	1.31	1.73	1.8				4.03	1.93
F10			1.56	3.46		1.25	9.23				
F31		3.93	2.14			1.71		4.03			
I87		2.57	2.14		2.79	1.79		1.93			

J96	Respiratory failure, not elsewhere classified
G47	Sleep disorders
E66	Overweight and obesity
J44	Other chronic obstructive pulmonary disease
I11	Hypertensive heart disease
J45	Asthma
F17	Nicotine dependence
E03	Other hypothyroidism
F10	Alcohol related disorders
F31	Bipolar disorder
I87	Other disorders of veins

part_Dec18D_cf1.html

Result cluster 3/10

CLUSTER 3

	Z96	H40	H35	H43	H34	H04	H25
Z96		7.49	6.76	9.3	6.62	5.29	4.59
H40	7.49		3.67		10.66	8.51	
H35	6.76	3.67		10.84			5.21
H43	9.3		10.84				
H34	6.62	10.66					
H04	5.29	8.51					
H25	4.59		5.21				

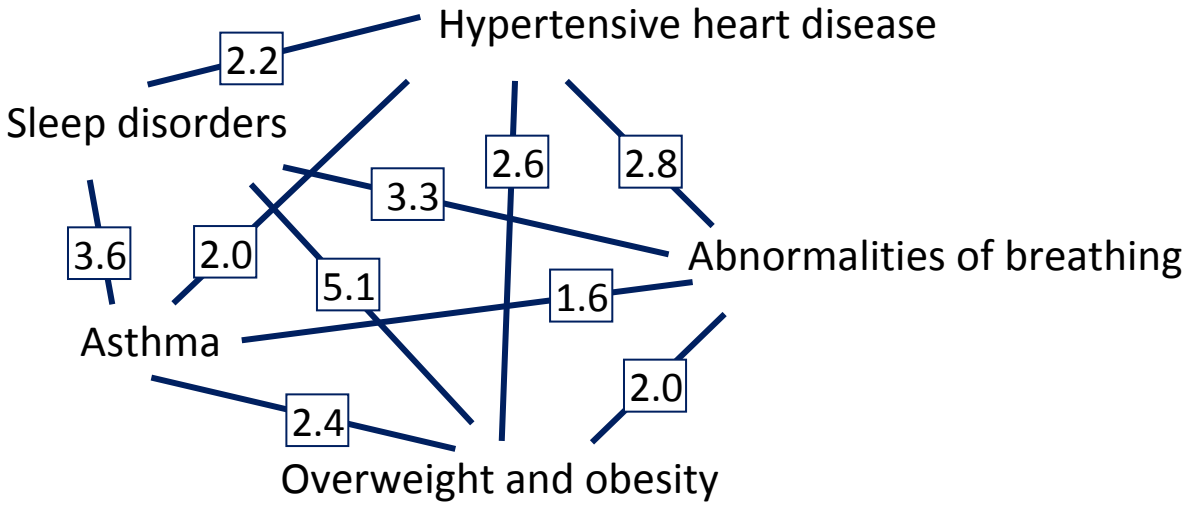
Z96 Presence of other functional implants
H40 Glaucoma
H35 Other retinal disorders
H43 Disorders of vitreous body
H34 Retinal vascular occlusions
H04 Disorders of lacrimal system
H25 Age-related cataract

Result cluster 1/10

CLUSTER 1

	I35	Z95	I50	I48	I25	I70	N18	M10	I36	I34	I21	I44	I42	K57	I71	I63	I69	F01	I49
I35		3.55	3.65	2.42	1.93	1.99	1.88	1.97	15.72	10.0		2.53		2.17	4.12		1.63		
Z95	3.55		2.83	2.28	3.35	2.28	1.66	2.17		3.86	3.09	6.53	3.52	1.93	3.03	1.59	1.74		5.28
I50	3.65	2.83		3.5	2.45	2.59	4.13	3.66		4.14	3.14	3.16	4.59	1.89		1.78	1.79	2.65	1.85
I48	2.42	2.28	3.5		1.78	1.8	2.16	2.4	6.0	3.78	1.61	2.07	2.34	1.62	1.66	2.38	2.2	1.99	3.29
I25	1.93	3.35	2.45	1.78		2.08	2.04	1.84	2.51	2.13	3.55	1.92	1.39	1.78	2.27	1.95	1.69	1.71	1.87
I70	1.99	2.28	2.59	1.8	2.08		2.4	1.63			3.38	1.89	1.3	1.49	3.19	2.32	1.93	1.51	
N18	1.88	1.66	4.13	2.16	2.04	2.4		4.87			3.37			2.53		2.47	2.01		
M10	1.97	2.17	3.66	2.4	1.84	1.63	4.87						3.55	2.37			1.31		
I36	15.72			6.0	2.51														
I34	10.0	3.86	4.14	3.78	2.13														
I21		3.09	3.14	1.61	3.55	3.38	3.37										1.52		
I44	2.53	6.53	3.16	2.07	1.92	1.89											1.53		
I42		3.52	4.59	2.34	1.39	1.3		3.55											
K57	2.17	1.93	1.89	1.62	1.78	1.49	2.53	2.37									1.57		
I71	4.12	3.03		1.66	2.27	3.19													
I63		1.59	1.78	2.38	1.95	2.32	2.47										7.05	6.14	
I69	1.63	1.74	1.79	2.2	1.69	1.93	2.01	1.31			1.52	1.53		1.57		7.05		4.6	1.82
F01			2.65	1.99	1.71	1.51										6.14	4.6		
I49		5.28	1.85	3.29	1.87												1.82		

I35	Nonrheumatic aortic valve disorders
Z95	Presence of cardiac and vascular...
I50	Heart failure
I48	Atrial fibrillation and flutter
I25	Chronic ischemic heart disease
I70	Atherosclerosis
N18	Chronic kidney disease (CKD)
M10	Gout
I36	Nonrheumatic tricuspid valve disorders
I34	Nonrheumatic mitral valve disorders
I21	Acute myocardial infarction
I44	Atrioventricular and left bundle-br...
I42	Cardiomyopathy
K57	Diverticular disease of intestine
I71	Aortic aneurysm and dissection
I63	Cerebral infarction
I69	Sequelae of cerebrovascular disease
F01	Vascular dementia
I49	Other cardiac arrhythmias

Conclusions

- Currently different diagnoses are treated separately
- Treating a related diseases together can be more cost effective
- In training new doctors, could optimize skill sets based on how diagnoses are connected