Subgrouping multimorbid patients with ischemic heart disease by means of unsupervised clustering: A cohort study of 72,249 patients defined comprehensively by diagnoses prior to presentation

Short title: Unsupervised clustering of patients with ischemic heart disease

Amalie D. Haue, PhD[1,2,¶], Peter C. Holm, MSc[1,¶], Karina Banasik, PhD[1], Agnete T. Lundgaard, PhD[1], Victorine P. Muse, MEng[1], Timo Röder, MSc[1], David Westergaard, PhD[1], Piotr J. Chmura, MSc[1], Alex H. Christensen, PhD[2,3], Peter E. Weeke, PhD[2], Erik Sørensen, PhD[4], Ole B. V. Pedersen, PhD[4,5], Sisse R. Ostrowski, DMSc[4,6], Kasper K. Iversen, DMSc[3], Lars V. Køber, DMSc[2,6], Henrik Ullum, DMSc[7], Henning Bundgaard, DMSc[2,5*], Søren Brunak, PhD[1,8*]

[1]Novo Nordisk Foundation Center for Protein Research, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark
[2]Department of Cardiology, The Heart Center, Rigshospitalet, Copenhagen, Denmark
[3]Department of Cardiology, Copenhagen University Hospital, Herlev, Denmark.
[4]Department of Clinical Immunology, Copenhagen University Hospital, Copenhagen, Denmark
[4]Department of Clinical Immunology, Copenhagen University Hospital, Copenhagen, Denmark
[5]Department of Clinical Immunology, Zealand University Hospital, Køge, Denmark
[6]Department of Clinical Medicine, University of Copenhagen, Rigshospitalet, Copenhagen, Denmark
[7]Statens Serum Institut, Copenhagen, Denmark
[8]Copenhagen University Hospital, Rigshospitalet, Copenhagen, Denmark

[*]E-mail: soren.brunak@cpr.ku.dk (SB)

¶These authors contributed equally to this work.

Total word count (including Title Page, Abstract, Text, References, Tables and Figures Legends: 7,589

Abstract

Background: There are no methods for classifying patients with ischemic heart disease (IHD) based on the entire spectrum of pre-existing diseases. Such methods might be clinically useful due to the marked differences in presentation and course of disease.

Methods: A population-based cohort study from a Danish secondary care setting of patients with IHD (2004-2016) and subjected to a coronary angiography (CAG) or coronary computed tomography angiography (CCTA). Data sources were The Danish National Patient Registry, in-hospital laboratory data, and genetic data from Copenhagen Hospital Biobank. Comorbidities included diagnoses assigned prior to presentation of IHD. Patients were clustered by means of the Markov Clustering Algorithm using the entire spectrum of registered multimorbidity. The two prespecified outcomes were: New ischemic events (including death from IHD causes) and death from non-IHD causes. Patients were followed from date of CAG/CCTA until one of the two outcomes occurred or end of follow-up, whichever came first. Biological and clinical appropriateness of clusters was assessed by comparing risks (estimated from Cox proportional hazard models) in clusters and by phenotypic and genetic enrichment analyses, respectively.

Findings: In a cohort of 72,249 patients with IHD (mean age 63.9 years, 63.1% males), 31 distinct clusters (C1-31, 67,136 patients) were identified. Comparing each cluster to the 30 others, seven clusters (9,590 patients) had statistically significantly higher or lower risk of new ischemic events (five and two clusters, respectively). 18 clusters (35,982 patients) had a higher or lower risk of death from non-IHD causes (12 and six clusters, respectively). All clusters at increased risk of new ischemic events, associated with risk of death from non-IHD causes as well. Cardiovascular or inflammatory diseases were commonly enriched in clusters (13), and distributions for 24 laboratory test results differed significantly across clusters.

Clusters enriched for cerebrovascular diseases were generally not at increased risk of the two outcomes. Polygenic risk scores were increased in a total of 15 clusters (48.4%).

Conclusions: Clustering of patients with IHD based on pre-existing comorbidities identified subgroups of patients with significantly different clinical outcomes and presented a tool to rank pre-existing comorbidities based on their association with clinical outcomes. This novel method may support better classification of patients and thereby differentiation of treatment intensity depending on expected outcomes in subgroups.

Non-standard abbreviations

CAG: Coronary arteriography

CCTA: Coronary computed tomography angiography

ICD-10: International Statistical Classification of Diseases and Related Health Problems 10th Revision

IHD: Ischemic heart disease

MCL: Markov clustering

NPR: Danish National Patient Registry

O/E-ratio: Observed-expected-ratio

PRS: Polygenic risk score

Introduction

Ischemic heart disease (IHD) is a common, chronic, and complex disease and mode of onset, disease burden and disease progression vary considerably between patients[1–3]. This heterogeneity relates to several factors, but a major contribution is multimorbidity as more than 85% of IHD patients have been diagnosed with other chronic diseases; a phenomenon coined cardiometabolic multimorbidity[4,5]. The increased mortality in patients with cardiometabolic multimorbidity is generally only related to single disease states, such as obstructive lung disease, diabetes, or stroke, although it is known that the risk of cardiovascular diseases is increased in many chronic, inflammatory disorders[6,7]. As more patients at older age and with more and more co-morbidities are seen, new methods for characterizing and studying cardiometabolic multimorbidity are needed[8–12].

Unsupervised clustering algorithms and other network-based methods can systematically reveal structures in large, feature-rich datasets and may be used to identify distinct patient subgroups within a heterogenous population[13,14]. Proof-of-concept analyses of cardiovascular phenotypes, including IHD, heart failure, diabetes, and atrial fibrillation have already been performed[15–21]. While these studies successfully identify subgroups resembling those from traditional analyses, they often fail to demonstrate that clustering analysis leads to novel understanding of a given dataset. Rather, they are typically restricted to characterize high-, medium-, and low-risk subgroups which by and large resemble more conservative approaches from an earlier, less data-rich, epoch[22].

For decades, Danish healthcare registries have had a strong position within epidemiological research[22–24]. Given the opportunities for using clinical data more extensively, we carried out an unsupervised clustering analysis of 72,249 patients with IHD based on their entire disease

history until IHD onset. Explicitly, we wanted to classify IHD based on the entire spectrum of multimorbidity. We identified distinct patient subgroups derived from a pool of 3,046 different diagnoses assigned prior to IHD onset. The biological and clinical factors characteristic to distinct patient subgroups identified by unsupervised clustering analysis were asserted by assessments of their associations with clinical outcomes and clinical characteristics, laboratory data, and genetics (Figure 1).

Methods

Data sources, study population, and outcomes

Data from the Danish National Patient Registry (NPR) and the Danish Registry for Causes of Death were linked to in-hospital electronic health data covering the two Danish healthcare regions in Eastern Denmark (~2.9 mil inhabitants), and the Copenhagen Hospital Biobank Cardiovascular Disease Cohort[23,25,26]. Linkage of different healthcare data sources was obtained via the personal identification number and only patients admitted to a hospital in Eastern Denmark in years 2004 to 2016 were considered[27]. We identified all patients in NPR who were assigned an ICD-10 code for IHD[28]. To increase the positive predictive value of IHD diagnoses and align included patients in time, we further required that patients had been subjected to coronary arteriography (CAG) or coronary computed tomography angiography (CCTA). To qualify that CAG/CCTAs were conclusive for IHD, patients were only included if the CAG/CCTA was performed during a contact where patients were assigned an ICD-10 code for IHD. We set the earliest CAG/CCTA fulfilling this criterium as the index date and excluded patients with an index date before year 2004 or after 2016 (Fig 2).

There were two predefined outcomes: 1) New ischemic events and 2) Death from other causes than IHD (non-IHD causes). The outcome “new ischemic event” was a composite outcome of a) hospitalization minimum 30 days after index for myocardial infarction or

unstable angina pectoris (i.e., hospitalization with myocardial infarction or unstable angina pectoris as the primary diagnosis), b) revascularization not related to the index date, and c) any death where IHD was listed as the primary or secondary cause. Outcomes were obtained from NPR and Danish Registry for Causes of Death. Eligible codes for inclusion, outcomes and specific cutoffs are available in S1 Fig and S1 Table.

Data preprocessing and application of the Markov cluster algorithm

We performed a clustering analysis of included patients based on their multimorbidity prior to their IHD diagnosis (index) using the Markov cluster (MCL) algorithm[29]. Multimorbidity was represented as patient-specific vectors using diagnoses assigned prior to or at index. ICD-10 codes assigned to less than five patients (n=1,673) were excluded from the analysis. As we focused the studies on multimorbidity in IHD, ICD-10 codes for IHD (I20-I25) were excluded from patients-specific vectors. Thus, a total of 3,046 ICD-10 codes were the basis for constructing a patient similarity network that was used as MCL algorithm input. Patient-specific vectors of length 3,046 with integers indicating the number of times a patient had been assigned a particular ICD-10 code. The length of the vectors corresponded to the number of input features (ICD-10 codes). By combining the patient-specific vectors from all included patients, a matrix of size n x m was constructed, where n indicates the number of included patients and m indicates the number of input features (ICD-10 codes). Following a series of preprocessing steps described in S1 Appendix, a patient similarity network was created based on the n x m matrix and used as input for the MCL algorithm[30]. Resulting clusters were denoted *C* followed by an integer indicating the rank of the clusters with respect to cluster size (number of patients in that cluster). Thus, C1 denotes the largest cluster and cluster-membership was used to denote a cluster as a covariate in subsequent analyses. Robustness of clustering was assessed by generating a series of diluted and shuffled versions

147 of the resulting clusters (reference clustering), and their similarity was quantified using the
148 variance of the information measure as previously described[31]. Explicitly, a series of diluted
149 and shuffled versions of the input graph were generated[32]. In total, 20 variations of the input
150 graph were constructed by shuffling and deleting edges, respectively. The variation in the
151 graphs was then quantified by means of variation of the information meassure. Details
152 regarding the MCL settings and a description of cluster robustness assessment are available
153 in the S1 Appendix.
154
155 *Preprocessing of laboratory and genetic data*
156 Clusters were characterized by laboratory and genetic data based on the subset of patients
157 where these data types were available. A panel of 25 different lab parameters was included in
158 the analyses. Only tests taken up to 90 days before index or at the day of index were included.
159 Included lab tests were plasma levels of potassium, sodium, hemoglobin, estimated
160 glomerular filtration rate (eGFR), creatinine, carbamide, glucose, troponin (I/T), HDL
161 cholesterol, LDL cholesterol, total cholesterol, leukocytes, C-reactive protein, lymphocytes,
162 monocytes, neutrophils, basophiles, platelets, INR, alanine transaminase, albumin, alkaline
163 phosphatase, bilirubin, and triglyceride. For every cluster, a *score* was computed based on the
164 number of patients with a lab test below, within, or above the standard reference value,
165 indicated by -1, 0 and 1, respectively. The *score* was defined as the mean of the summarized
166 values per cluster.
167
168 Autosomal genotype data were obtained by identifying included patients who were also
169 among the study participants in the Copenhagen Hospital Biobank – Cardiovascular Disease
170 Cohort[26]. For included patients with genetic data available, we calculated polygenic risk
171 scores (PRSs) for 14 traits, obtained from nine GWAS meta-analyses (atrial fibrillation, BMI-

adjusted non-insulin diabetes, chronic kidney disease, HDL cholesterol levels, heart failure, LDL cholesterol levels, stroke, total cholesterol levels, triglyceride levels) and five GWAS (acute myocardial infarction, coronary artery disease, diastolic blood pressure, non-alcoholic fatty liver disease, systolic blood pressure)[39–42]. PRSs were calculated using the "LDpred2-auto" algorithm, implemented in the R package "bigsnpr" (version 1.11.6) with R version 4.0.0 and the workflow management system Snakemake[43–45]. Each trait's PRS distribution was scaled to a mean of zero and a standard deviation of one.

Statistical analyses of clusters identified by the MCL algorithm

As the study was designed to identify patient subgroups and not individual variation, clusters of size < 500 were excluded from the remaining analyses. Mean age at IHD onset in each cluster was compared to the mean age at onset in all the other clusters using Tukey's Honest Significant Difference (HSD) method. Significance level was set to 0.05 and P-values were adjusted using the Holm method assuming 465 tests (adj. P-val.).

To investigate the association between cluster-membership and the competing risks of new ischemic events and death from non-IHD causes, we used Cox proportional-hazards models (Cox models). Patients were followed from index until occurrence of either of the two outcomes, or end of follow-up (year 2018), whichever came first. The dependent variable was either risk of new ischemic events or death from non-IHD causes, and the independent variables were cluster, sex, and age at index. To age-adjust the models, analyses were performed using restricted cubic spline with three knots for age at index. Follow-up time was truncated to a maximum of five years. For each cluster, hazard ratios (HRs) and 95% confidence intervals (CIs) were estimated by comparing HRs for the members of the cluster with the HRs with that of non-members.

Further characterization of clusters consisted of: (1) phenotypic enrichment analysis, (2) characterization of clusters with respect to their laboratory profiles and (3) a test for genetic enrichment. The phenotypic enrichment analysis was carried out based on ratios between observed (O) and expected (E) frequencies of diagnoses in the clusters (O/E-ratios). That is, ratios between the frequencies of ICD-10 codes in each cluster (observed frequencies) and the frequencies of ICD-10 codes in the entire population (expected frequencies) were calculated and expressed as O/E-ratios[46]. In subsequent characterization of clusters, enrichment denoted O/E-ratios > 2, and clusters were characterized as having little enrichment if the sum of the ten largest O/E-ratios < 50. Inverse changes were used to denote O/E-ratios between 0 and 1.

Hierarchical clustering was applied to estimate the cluster similarity with respect to the laboratory tests using the Euclidean distance between the *score* of each cluster for each test.

For each of the fourteen traits we calculated PRSs for, we used Wilcoxon rank-sum tests to compare the PRS distribution of each cluster to the combined PRS distribution of PRSs in all other clusters. Resulting P-values were converted to the false discovery rate (FDR) to account for multiple testing, with a total of 434 tests. We report effect sizes as calculated by the "wilcox.test" function built into R version 4.0.0. Level of significance was set to FDR < 0.05, assuming 434 tests.

Further details regarding preprocessing and analyses of laboratory and genetic data are available in the S2 Appendix.

Ethics approvals and data access

The study was approved by The National Ethics Committee (1708829, 'Genetics of CVD'—a genome-wide association study on repository samples from Copenhagen Hospital Biobank), The Danish Data Protection Agency (ref: 514-0255/18-3000, 514-0254/18-3000, SUND-2016-50), The Danish Health Data Authority (ref: FSEID-00003724 and FSEID-00003092), and The Danish Patient Safety Authority (3-3013-1731/1/). Danish personal identification numbers were pseudonymized prior to any analysis. Study design, methods and results were reported in agreement with the STROBE statement[47].

Results

Cohort demographics and co-morbidities

A total of 72,249 patients (63.1% males, mean age 63.9 years) were included (Table 1). Angina pectoris (I20) was the most common IHD diagnosis (38,239 patients, 52.9%), followed by acute myocardial infarction (I21) (33,229 patients, 46.0 %) and chronic IHD (I25) (22,750 patients, 31.5%). The most common co-morbidity prior to the IHD index was hypertension (I10.9) (24,818 patients, 34.4%) followed by dyslipidemia (E78.0) (12,780 patients, 17.7%) and non-insulin dependent diabetes (E11.9) (7,551 patients, 10.5%). Prior to index, the mean number of diagnoses per patient was 8.1. A total of 68,103 patients (94.3%) had co-morbidities registered prior to index. The overall incidence (new ischemic events and death from non-IHD causes) was 94 events per 1000 person-years (Table 1).

Unsupervised clustering of multimorbid patients with IHD

In the cohort, the MCL algorithm identified 36 distinct clusters based on the set of 3,046 ICD-10 codes assigned to the patients prior to or at index. The 36 clusters contained a total of 68,084 patients. Expectedly, the remaining 4,365 patients (6.0% of included patients) that did not cluster were primarily patients with no diagnoses prior to index (>99%). Further, cluster

247 robustness was assessed as described in Methods, where the variation of information measure
248 less than 2 if 25% of the edges in the input graph were deleted or shuffled (S4 Figure). Next,
249 the 31 of the 36 clusters with >500 patients (67,136 patients) were characterized (Table 2).
250 Using Tukey's HSD to compare the age at index between all 31 clusters (a total of 466
251 combinations), we found significant differences in 391 comparisons (84.1%, S3 Table). For
252 demographics of patients that did not cluster or were in clusters of size < 500, see S4 Table.

Clusters, clinical outcomes, and phenotypic enrichment

To assess if the unsupervised clustering identified patient subgroups at different risks of disease progression, we used cluster-membership (C1-C31) as a covariate in a series of Cox models. A total of 14,679 patients experienced a new ischemic event during follow-up and 10,684 patients died from other causes than IHD. Mean follow-up time was 3.72 years (Table 1). Risks for new ischemic events and death from non-IHD causes in each cluster were compared to the pooled risk for patients in the remaining 30 clusters. The survival analysis demonstrated that the MCL algorithm stratified patients according to risk of new ischemic events and death from non-IHD causes (Fig 3). Comparing each cluster (n=1) to all the others (n=30), a total of seven clusters (20,221 patients) had a statistically significantly higher or lower risk of new ischemic events (Adj. P-val. < 0.05). Five clusters (9,590 patients) and two clusters (10,631 patients) were at increased and decreased risk of new ischemic events, respectively. Similarly, a total of 18 clusters (43,173 patients) had a statistically significantly higher or lower risk of death from non-IHD causes (Adj. P-val. < 0.05); where 14 clusters (21,282 patients) and four clusters (21,891 patients) were at increased or decreased risk of death from non-IHD causes. All clusters at increased risk of new ischemic events, associated with risk of death from non-IHD causes as well. The same was true for the two clusters at decreased risk of new ischemic events, i.e., these clusters were at decreased risk of death from non-IHD causes as well. A total of 13 clusters, (23,963 patients) were not have altered risk of the two outcomes, when compared to the other clusters (Table 2).

The distribution of O/E-ratios was heavily left-skewed as less than 99% (n=101) of all O/E-ratios were >10 and roughly 7% (n=887) of all O/E-ratios were >2. About 60% of all O/E-ratios (n=8,056) were in the range of 0 and 1 corresponding to inverse changes. Generally, clusters that had high risk of new ischemic events or death from non-IHD causes were also

characterized by large, summarized O/E-values corresponding to a high degree of multimorbidity (S5 Table 5). The results of the enrichment analysis were summarized according to nine different disease categories: (1) diabetes mellitus, (2) cardiac diseases, (3) diseases affecting the upper airways, (4) cerebrovascular diseases, (5) infections and other acquired diseases, (6) gynecologic diseases, (7) Inflammatory and degenerative of the musculoskeletal system, (8) diseases of the urinary system, and (9) hypertension (Fig. 4).

An in-depth characterization of clusters enriched for cardiometabolic or -vascular diseases, degenerative or inflammatory diseases and clusters characterized by little enrichment and inverse changes is provided in the following paragraphs.

Clusters enriched for cardiometabolic and -vascular diseases

Four of the five clusters at increased risk of new ischemic events (and death from non-IHD causes) were enriched for diabetes (C5, C18, C23, and C30). In these four clusters, HRs ranged from 1.40 (C5, 95%CI: 1.30;1.50, adj. P-val. < 0.001) to 1.88 (C30, 95%CI: 1.60;2.00, adj. P-val. < 0.001) with a significant difference in age at index (C5: 63.9 years, C30: 61.2 years, Adj. P-val. < 0.001, TukeyHSD). C18 and C23 were only enriched for insulin-dependent diabetes, but differed in that C18 was also enriched for insulin-dependent diabetes with vascular complications and periphery atherosclerosis. In contrast, C5 was only enriched for non-insulin dependent diabetes and included diabetes with as well as without complications. Lastly, C30 was only enriched for diabetes with complications (insulin and non-insulin dependent) and was the diabetes cluster enriched for chronic kidney disease and bacterial infections, as well (S5 Table 5).

Other cardiac diseases that displayed enrichment were supraventricular arrythmias (C4), cardiomyopathies (C9), and valve diseases (C20). Of the three clusters, only C9 had increased risk of new ischemic events (HR: 1.31 (C9, 95%CI: 1.20;1.44, Adj. P-val: < 0.001). Risk of death from non-IHD causes was 1.79 (95%CI: 1.60;2.00, adj. P-val. < 0.001). In contrast, C4 and C20 only had increased risk of death from non-IHD causes with HRs of 1.49 (C4, 95%CI: 1.34;1.59, adj. P-val. < 0.001) and 1.78 (C20, 95%CI: 1.54;2.04, adj. P-val. < 0.001). Interestingly, the cluster enriched for cerebrovascular diseases (C27) did not have altered risk of any of the two outcomes. In sum, all clusters that had increased risk of new ischemic events were enriched for cardiometabolic diseases, albeit not all clusters enriched for cardiometabolic and -vascular diseases had increased risk of new ischemic events (Table 2 and S5 Table 5).

Clusters enriched for degenerative or inflammatory diseases

Six clusters (C7, C13, C14, C22, C26, and C31) were enriched for diagnoses describing degenerative or inflammatory diseases, i.e., osteoarthritis (C7), degenerative spine disease (C13 and C22), chronic obstructive pulmonary disease (C14), asthma (C26), and rheumatoid arthritis (C31). Remarkably, none of the four clusters had increased risk of new ischemic events and only one cluster (C14) had increased risk of death from non-IHD causes (HR: 3.39, 95%CI: 3.09;3.71, adj. P-val. < 0.001). Conversely, C7 and C13 had reduced risk of death from non-IHD causes (C7, HR: 0.61, 95%CI: 0.52;0.72, adj. P-val. < 0.001 and C13, HR: 0.58, 95%CI: 0.45;0.74, adj. P-val. < 0.001). Age at index for the clusters enriched for degenerative or inflammatory diseases range between 58.6 years (C13) and 69.2 years (C22) (Table 2). Taken together, these findings hint to the dual nature of inflammation as a potential disease modifier as well as a risk factor.

Clusters characterized by little enrichment and inverse changes

Six clusters (C1, C2, C3, C6, C15, and C17) were characterized by little enrichment, which included the two clusters with reduced risk of new ischemic events (C2, HR: 0.82, 95%CI: 0.76;0.89, adj. P-val. < 0.001 and C3, HR: 0.76, 95%CI: 0.52;0.69, adj. P-val. < 0.001). Not surprisingly, none of these six clusters had increased risk of either of the two outcomes, but three clusters (C2, C3, and C6) had reduced risk of death from non-IHD causes (C2, HR: 0.60, 95%CI: 0.52;0.69, adj. P-val. < 0.001, C3, HR: 0.59, 95%CI: 0.59;0.69, adj. P-val. < 0.001 and C6, HR: 0.68, 95%CI: 0.57;0.79, adj. P-val. < 0.001) (Table 2). It was a common attribute of the clusters without altered risk of any of the two outcomes that O/E-ratios for hypertension and dyslipidemia were among the largest. In contrast, diabetes, heart failure, and chronic obstructive pulmonary disease frequently displayed inverse changes (O/E-ratios < 1) in these clusters (S5 Table). Taken together, these observations indicate that risk of disease progression in this populations necessitates a more sophisticated analysis of multimorbidity.

For a list with results of the enrichment analysis for all clusters, including the 13 clusters not described above, S5 Table 5.

Clusters and their association with laboratory measurements and genetic data

Clusters were also characterized by means of datatypes not included among the MCL algorithm input features. For patients in the 31 clusters, we had laboratory measurements on 30,755 (49.5%) and genetic data on 19,422 (31.3%). To assess if the phenotypic differences captured by the MCL algorithm were also reflected in laboratory measurements, we tested if the distributions of test results within and out of reference ranges differed significantly. There were significantly different distributions of tests within and out of reference ranges in clusters

for the 24 most frequent tests. Overall, this indicates that the phenotypic patterns within the entire spectrum of cardiovascular multimorbidity registered before index correlate with results of clinical laboratory tests (S6 Table). Thus, these findings are a strong indicator that the patterns captured by the MCL algorithm are biologically relevant. For a graphical summary of the laboratory scores in each cluster, see S5 Figure.

Finally, we identified 41 cases (out of 434 tests) where the PRS distribution for a specific trait in a cluster was significantly different from that trait's combined PRS distribution of the other 30 clusters. Among these cases, we found the largest effects size to be a higher genetic risk for atrial fibrillation in cluster C4 (0.57, FDR < 0.001) as well as a higher genetic risk for non-insulin dependent diabetes in cluster C5 (0.55, FDR < 0.001). These findings are congruent with the results of the enrichment analysis for C4 and C5, respectively. In contrast, C1 (phenotypically characterized by inverse changes) had relatively large, positive effect sizes for systolic as well as diastolic blood pressure (0.20 and 0.16, FDR < 0.001). Similarly, there were positive effect sizes for total cholesterol and triglycerides in C6, which was also characterized by little phenotypic enrichment as well as a high degree of inverse changes. A list of significant effect sizes for the 41 significant cases, see S7 Table.

Discussion

In this study, we developed a novel, data-driven method for structuring the entire spectrum of multimorbidity by means of an unsupervised clustering analysis. In a cohort of 72,249 patients with IHD patients, we identified 31 distinct clusters (67,136 patients) based on 3,046 diagnoses assigned prior to or at index. By comparing risk of new ischemic events and death from non-IHD causes across clusters and then performing an enrichment analysis, we found that clusters at increased risk of new ischemic events were enriched for diabetes (four clusters) or cardiomyopathies (one cluster). Neither the cluster enriched for supraventricular

arrythmias, nor valve diseases had increased risk of new ischemic events. Degenerative and inflammatory diseases were enriched in a total of six clusters and displayed no clear trend in their relation to the outcomes. The results of the enrichment analysis were supported by trends in laboratory test results and clusters enriched for supraventricular arrythmias and non-insulin diabetes also had congruently, higher genetic risks.

The results of the study agree with common knowledge on risk of IHD, while also adding insights to the disease-diseases associations, which are currently underappreciated in the literature. The fact that clusters enriched for diabetes were generally the most high-risk clusters serves as a methodological reality check[6]. Added value of the study lies in the fact that the method allows for a more sophisticated description of such associations, as the method allows to study the entire spectrum of multimorbidity. For example, four clusters were enriched for diabetes, which is in line with the current paradigm that a single term is insufficient to describe a multifactorial disease, such as diabetes[18,31]. By integrating different data types, the findings indicate how phenotypic and genetic data complement each other, by exemplifying (1) that clustering analysis facilitates stronger genetic signals in patient subgroups and (2) that genetic data may unveil patterns not captured by phenotypic data alone.

In addition, the method developed in this study and subsequent findings add perspective to the relatively limited body of literature regarding associations between chronic inflammatory and cardiovascular diseases[7]. While previous studies have concluded that the risk of cardiovascular diseases is increased in most chronic inflammatory disorders, the results of our study indicate that pre-existing degenerative or inflammatory disorders in patients with IHD do not increase the risk of new ischemic events.

The pre-selected outcomes in the present study are also a unique aspect of the study, as previous clustering analyses within the cardiovascular domain studies have mainly analyzed all-cause mortality[19,20]. This aspect of the study allows to distinguish between risk of progression related to IHD and risk of progression that is related to comorbidity drawing attention to important aspects of multimorbidity in this domain. For example, clusters enriched for supraventricular arrythmias and chronic obstructive pulmonary disease, respectively, only had increased risk of death from non-IHD causes. The study design, including the enrichment analysis, also revealed that classical risk factors for IHD (e.g., hypertension and dyslipidemia) did not drive the clustering. This finding agrees with previously published comorbidity phenotypes in patients with IHD[20]. We argue that the present study displays that continuous exploration and characterization of multimorbidity in IHD are key elements in optimizing the exploit the full potential of continuously developing treatment strategies.

Previous clustering analyses within the cardiovascular domain have typically included either thousands of patients or hundreds of input features, but not both[16,17]. For example, Hall et al. defined multimorbidity using only eight different chronic conditions, whereas Crowe et al. defined multimorbidity with reference to 20 predefined conditions[19,20]. Thus, the scale of our study exceeds that of previous work, as it includes more than 70,000 patients and more than 3,000 input features. And further, we limited the risk of introducing bias by not exerting feature selection prior to clustering.

The two main limitations with respect to the data foundation are that (1) owing to the novelty of the method, there were no standardized way of assessing the representation of multimorbidity and (2) it was only a subset for which laboratory and genetic data were

available. These challenges are naturally overcome in clustering analyses based on data from randomized controlled trials, such as the studies by Inohara et al, and Karwath et al.[17,21] However, in the present, data-rich era, we argue that it is highly important to develop methods for structuring and studying other data than what is being collected for trials. Ideally, the two approaches, based on nationwide data and randomized controlled trials, respectively, will complement each other; and will facilitate more precise identification of patients who are likely to benefit from different treatment options as well as guide optimized selection of patients for randomized controlled trials.

In conclusion, the study further showcases the strengths of a more fine-grained analysis of patient subgroups, which, in turn, may pave the way for successful implementation of precision medicine. Owing to its flexibility, the comprehensive, data-driven analysis of cardiovascular multimorbidity represents a novel method for characterizing multimorbidity in IHD with great potential of applying it to other diseases of interest or other clinical data. Such trends may guide clinical decision making in cases, where for example it is not obvious how to manage the angiographic findings or the combination of drugs that a specific patient will benefit most from.

In conclusion, the present study cements the complexity of multimorbid patients with IHD and exemplifies the clinical relevance of a more fine-grained patient subgrouping by carrying out a cluster-based risk-stratifying the cohort. Further, owing to its flexibility, the comprehensive, data-driven method of cardiovascular multimorbidity presented here represents a novel method for characterizing multimorbidity in IHD with great potential. Improved patient subgrouping may be critical guide future clinical decision making in cases,

where it is non-trivial how to manage the angiographic findings or to find the optimal combination of drugs for a given patient.

Funding

This work was financially supported by Novo Nordisk Foundation (Grants NNF17OC0027594 and NNF14CC0001) and the Innovation Fund Denmark via the NordForsk project PM Heart (5184-00102B).

Acknowledgement

The authors would like to thank (1) research programmer, Troels Siggaard, Novo Nordisk Foundation Center for Research, University of Copenhagen, Denmark for continuous and reliable infrastructure support, and (2) Head of Cardiovascular Research, Hilma Hólm, deCODE genetics, Iceland for insightful comments

Data access

Application for registry data access can be made to the Danish Health Data Authority (contact: servicedesk@sundhedsdata.dk). Anyone wishing access to the data and use them for research will be required to meet research credentialing requirements as outlined at the authority's web site: sundhedsdatastyrelsen.dk/da/english/health_data_and_registers/research_services. Requests are normally processed within three to six months.

Code availability statement

The code used to generate the results including the clustering pipeline will be made publicly available upon publication.

References

1. Antman, E. M. & Braunwald, E. Managing Stable Ischemic Heart Disease. *N. Engl. J. Med.* **382**, 1468–1470 (2020).
2. Ferraro, R. *et al.* Evaluation and Management of Patients With Stable Angina: Beyond the Ischemia Paradigm. *J. Am. Coll. Cardiol.* **76**, 2252–2266 (2020).
3. Nabel, E. G. & Braunwald, E. A tale of coronary artery disease and myocardial infarction. *N. Engl. J. Med.* **366**, 54–63 (2012).
4. Forman, D. E. *et al.* Multimorbidity in Older Adults With Cardiovascular Disease. *J. Am. Coll. Cardiol.* **71**, 2149–2161 (2018).
5. Afilalo, J. *et al.* Frailty Assessment in the Cardiovascular Care of Older Adults. *J. Am. Coll. Cardiol.* **63**, 747–762 (2014).
6. The Emerging Risk Factors Collaboration. Association of Cardiometabolic Multimorbidity With Mortality. *JAMA* **314**, 52–60 (2015).
7. Dregan, A., Charlton, J., Chowienczyk, P. & Gulliford, M. C. Chronic Inflammatory Disorders and Risk of Type 2 Diabetes Mellitus, Coronary Heart Disease, and Stroke. *Circulation* **130**, 837–844 (2014).
8. Glynn, L. G. Multimorbidity: another key issue for cardiovascular medicine. *The Lancet* **374**, 1421–1422 (2009).
9. Joshi, A., Rienks, M., Theofilatos, K. & Mayr, M. Systems biology in cardiovascular disease: a multiomics approach. *Nat. Rev. Cardiol.* **18**, 313–330 (2021).
10. Khera Amit V. & Kathiresan Sekar. Is Coronary Atherosclerosis One Disease or Many? *Circulation* **135**, 1005–1007 (2017).
11. Rahimi, K., Lam, C. S. P. & Steinhubl, S. Cardiovascular disease and multimorbidity: A call for interdisciplinary research and personalized cardiovascular care. *PLOS Med.* **15**, e1002545 (2018).

12. Haue, A. D. *et al.* Temporal patterns of multi-morbidity in 570157 ischemic heart disease patients: a nationwide cohort study. *Cardiovasc. Diabetol.* **21**, 87 (2022).
13. Kaufman, L. & Rousseeuw, P. J. *Finding Groups in Data: An Introduction to Cluster Analysis*. (John Wiley & Sons, 2009).
14. Lee, L. Y., Pandey, A. K., Maron, B. A. & Loscalzo, J. Network medicine in Cardiovascular Research. *Cardiovasc. Res.* **117**, 2186–2202 (2021).
15. Shah, R. V. *et al.* Association of Multiorgan Computed Tomographic Phenomap With Adverse Cardiovascular Health Outcomes: The Framingham Heart Study. *JAMA Cardiol.* **2**, 1236–1246 (2017).
16. Ahmad, T. *et al.* Clinical implications of chronic heart failure phenotypes defined by cluster analysis. *J. Am. Coll. Cardiol.* **64**, 1765–1774 (2014).
17. Inohara, T. *et al.* Association of of Atrial Fibrillation Clinical Phenotypes With Treatment Patterns and Outcomes: A Multicenter Registry Study. *JAMA Cardiol.* **3**, 54–63 (2018).
18. Ahlqvist, E. *et al.* Novel subgroups of adult-onset diabetes and their association with outcomes: a data-driven cluster analysis of six variables. *Lancet Diabetes Endocrinol.* **0**, (2018).
19. Hall, M. *et al.* Multimorbidity and survival for patients with acute myocardial infarction in England and Wales: Latent class analysis of a nationwide population-based cohort. *PLOS Med.* **15**, e1002501 (2018).
20. Crowe, F. *et al.* Comorbidity phenotypes and risk of mortality in patients with ischaemic heart disease in the UK. *Heart* **106**, 810–816 (2020).
21. Karwath, A. *et al.* Redefining β-blocker response in heart failure patients with sinus rhythm and atrial fibrillation: a machine learning cluster analysis. *The Lancet* **398**, 1427–1435 (2021).

22. Bowman, L. *et al.* Understanding the use of observational and randomized data in cardiovascular medicine. *Eur. Heart J.* **41**, 2571–2578 (2020).
23. Schmidt, M. *et al.* The Danish health care system and epidemiological research: from health care contacts to database records. *Clin. Epidemiol.* **11**, 563–591 (2019).
24. Hemingway, H. *et al.* Big data from electronic health records for early and late translational cardiovascular research: challenges and potential. *Eur. Heart J.* **39**, 1481–1495 (2018).
25. Helweg-Larsen, K. The Danish Register of Causes of Death. *Scand. J. Public Health* **39**, 26–29 (2011).
26. Sørensen, E. *et al.* Data Resource Profile: The Copenhagen Hospital Biobank (CHB). *Int. J. Epidemiol.* (2020) doi:10.1093/ije/dyaa157.
27. Schmidt, M., Pedersen, L. & Sørensen, H. T. The Danish Civil Registration System as a tool in epidemiology. *Eur. J. Epidemiol.* **29**, 541–549 (2014).
28. Sundbøll, J. *et al.* Positive predictive value of cardiovascular diagnoses in the Danish National Patient Registry: a validation study. *BMJ Open* **6**, (2016).
29. Enright, A. J., Van Dongen, S. & Ouzounis, C. A. An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Res.* **30**, 1575–84 (2002).
30. MCL - a cluster algorithm for graphs. http://micans.org/mcl/.
31. Kirk, I. K. *et al.* Linking glycemic dysregulation in diabetes to symptoms, comorbidities, and genetics through EHR data mining. *eLife* **8**, e44941 (2019).
32. Meilă, M. Comparing clusterings—an information based distance. *J. Multivar. Anal.* **98**, 873–895 (2007).
33. Nielsen, J. B. *et al.* Biobank-driven genomic discovery yields new insight into atrial fibrillation biology. *Nat. Genet.* **50**, 1234–1239 (2018).

34. Mahajan, A. *et al.* Fine-mapping type 2 diabetes loci to single-variant resolution using high-density imputation and islet-specific epigenome maps. *Nat. Genet.* **50**, 1505–1513 (2018).
35. Wuttke, M. *et al.* A catalog of genetic loci associated with kidney function from analyses of a million individuals. *Nat. Genet.* **51**, 957–972 (2019).
36. Surakka, I. *et al.* The impact of low-frequency and rare variants on lipid levels. *Nat. Genet.* **47**, 589–597 (2015).
37. Shah, S. *et al.* Genome-wide association and Mendelian randomisation analysis provide insights into the pathogenesis of heart failure. *Nat. Commun.* **11**, 163 (2020).
38. Malik, R. *et al.* Multiancestry genome-wide association study of 520,000 subjects identifies 32 loci associated with stroke and stroke subtypes. *Nat. Genet.* **50**, 524–537 (2018).
39. Jiang, L., Zheng, Z., Fang, H. & Yang, J. A generalized linear mixed model association tool for biobank-scale data. *Nat. Genet.* **53**, 1616–1621 (2021).
40. van der Harst, P. & Verweij, N. Identification of 64 Novel Genetic Loci Provides an Expanded View on the Genetic Architecture of Coronary Artery Disease. *Circ. Res.* **122**, 433–443 (2018).
41. Hoffmann, T. J. *et al.* Genome-wide association analyses using electronic health records identify new loci influencing blood pressure variation. *Nat. Genet.* **49**, 54–64 (2017).
42. Anstee, Q. M. *et al.* Genome-wide association study of non-alcoholic fatty liver and steatohepatitis in a histologically characterised cohort☆. *J. Hepatol.* **73**, 505–515 (2020).
43. Privé, F., Arbel, J. & Vilhjálmsson, B. J. LDpred2: better, faster, stronger. *Bioinformatics* **36**, 5424–5431 (2020).
44. R Core Team. *R: A Language and Environment for Statistical Computing*. (R Foundation for Statistical Computing, 2020).

45. Mölder, F. *et al.* Sustainable data analysis with Snakemake. Preprint at https://doi.org/10.12688/f1000research.29032.2 (2021).

46. Violán, C. *et al.* Multimorbidity patterns with K-means nonhierarchical cluster analysis. *BMC Fam. Pract.* **19**, 108 (2018).

47. Elm, E. von *et al.* The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *J. Clin. Epidemiol.* **61**, 344–349 (2008).

Figure legends

Fig 1: Graphical overview of study. Conceptual figure displaying the study design. A: Assemblage of patient-specific vectors that were the basis for construction of a matrix and an *n* x *m* matrix, where *n* corresponds to the number of included patients and *m* corresponds to the number of diagnoses. B: Unsupervised clustering of IHD patients using the MCL algorithm, which was the basis for performing unsupervised clustering to identify distinct clusters and associating them with clinical outcomes. C: Risk of disease progression (new ischemic events or death from non-IHD causes) in clusters. Color bar indicates increased, not altered, or decreased risk for patients in one cluster relative to the patients not in that cluster. D: Phenotypic and genetic characterization of clusters. Red: Increased risk of both outcomes. IHD: Ischemic heart disease. MCL: Markov Clustering.

Fig 2: Flowchart: Data sources, study population, and outcomes. Gray: Identification. Blue: Screening. Red: Eligibility. Green: Inclusion and outcomes. AMI: Acute myocardial infarction. UAP: Unstable angina pectoris. NPR: The Danish National Patient Registry. IHD: ischemic heart disease (ICD-10 codes I20-I25). CAG: Coronary arteriography. CCTA: Coronary computed tomography angiography. ICD-10: International Statistical Classification of Diseases and Related Health Problems 10th Revision. SKS: Sundhedsvæsenets Klassifikationssystem (The Danish Health Authority Classification System).

Fig 3: Risk of new ischemic events and non-IHD causes stratified by cluster. Forest plots where clusters are shown against HR for new ischemic events (left) and death from non-IHD causes (right). X-axis: HR for a single cluster relative to mean HR of the 30 other clusters. Y-axis: Clusters arranged by risk of new ischemic events, increasing risk from top to bottom. Colors indicating significance. Dark green: Reduced risk of new ischemic events and

death from non-IHD causes. Lighter green: Reduced risk of death from non-IHD causes. Yellow: No significance. Orange: Increased risk of death from non-IHD causes. Red: Increased risk of new ischemic events and increased risk of death from non-IHD causes. IHD: Ischemic heart disease. HR: Hazzard ratio.

Fig 4: Infographic summarizing the results of the study. Center: Study cohort. Periphery: Graphical overview of results from clustering analysis, survival analysis and characterization of clusters. Arrows indicate disease categories (for details, see text). 1: Diabetes mellitus. 2: Cardiac diseases. 3: Diseases affecting the upper airways. 4: Cerebrovascular diseases. 5: Infections and other acquired diseases. 6: Gynecologic diseases. 7: Inflammatory and degenerative of the musculoskeletal system. 8: Diseases of the urinary system. 9: Hypertension. C1-31: Clusters. “Underline” indicates little enrichment. “*” indicates genetic enrichment. For underlying data, see S5 and S7 Tables.

Table 1: Patient demographics, co-morbidities, and outcomes

Cohort demographics	**Total**	**Males**	**Females**
Number of patients (%)	72,249	45,576 (63.1)	26,673 (36,1)
Mean age at index (SD)	63.9 (11.9)	62.9 (11.6)	65.6 (12.1)
IHD manifestations (ICD-10)	**Total**	**Males**	**Females**
Angina pectoris (I20)	38,239	22,628	15,611
Acute myocardial infarction (I21)	33,299	27,720	10,579
Subsequent myocardial infarction (I22)	61	34	27
Certain current complications following acute myocardial infarction (I23)	138	92	46
Other acute ischemic heart diseases (I24)	1,341	814	527
Chronic ischemic heart disease (I25)	22,750	14,589	8,152
Common comorbidities (ICD-10)	**Total**	**Males**	**Females**
Primary (essential) hypertension (I10.9)	24,818	14,508	10,310
Hypercholesterolemia (E78.0)	12,780	7,842	4,938
Non-insulin dependent diabetes (E11.9)	7,551	4,891	2,660
Atrial fibrillation and atrial flutter, unspecified (I48.9)	7,075	4,509	2,566
Heart failure, unspecified (I50.9)	6,160	4,059	2,101
Chest pain, unspecified (R07.9)	5,863	3,441	2,422
Senile cataract, unspecified (H25.9)	5,764	2,795	2,969
Pneumonia, unspecified (J18.9)	5,469	3,236	2,260
Hyperlipidaemia, unspecified (E78.5)	5,002	3,306	1,696
Chronic obstructive pulmonary disease (J44.9)	4,621	2,449	2,172
Outcomes, number of cases	**Total**	**Males**	**Females**
New ischemic events (%)	14,679	10,152	4,527
■ *Myocardial infarction*	*5,833*	*3,709*	*2,124*
■ *Revascularization*	*6,282*	*4.718*	*2,124*
■ *Death caused by IHD*	*2,563*	*1.724*	*839*
Death from non-IHD causes (%)	10,684	6,710	3,974
Censored (%)	46,886	28,713	18,172
Outcomes, time to event	**Mean time to event in years (SD)**		
	Total	**Males**	**Females**
New ischemic events	1.48 (1.40)	1.49 (1.41)	1.48 (1.40)
■ Myocardial infarction	2.40 (1.87)	2.41 (1.89)	2.38 (1.85)
■ Revascularization	2.25 (1.88)	2.28 (1.89)	2.16 (1.84)
■ Death caused by IHD	1.92 (1.13)	1.95 (2.02)	1.88 (2.05)
Death from non-IHD causes	2.16 (1.50)	2.14 (1.49)	2.20 (1.51)
Censored	4.37 (1.08)	4.36 (1.09)	4.39 (1.06)
Total	3.72 (1.64)	3.67 (1.67)	3.81 (1.60)

618

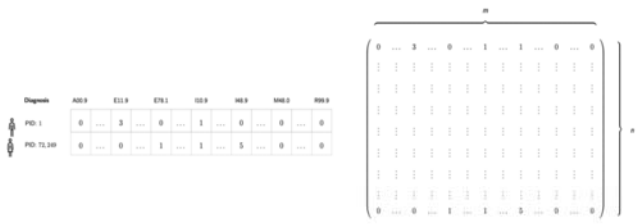
Table 2: Cluster demographics, characteristics, and associations with outcomes

Cluster	Size	Mean age at index in years (SD)	Males	Females	New ischemic events		Death from non-IHD causes	
					HR	Adj. P-val.	HR	Adj. P-val.
C1	7,191	64.8 (11.3)	3,897	3,294	1.000	> 0.050	0.856	> 0.050
C2	5,990	58.6 (11.5)	2,862	3,127	0.825	**< 0.001**	0.600	**< 0.001**
C3	4,641	56.8 (11.4)	2,727	1,914	0.757	**< 0.001**	0.586	**< 0.001**
C4	4,401	69.6 (10.2)	2,853	1,548	0.920	> 0.050	1.461	**< 0.001**
C5	4,290	63.9 (10.7)	2,803	1,487	1.402	**< 0.001**	1.629	**< 0.001**
C6	3,589	59.7 (10.9)	2.388	1,201	0.969	> 0.050	0.675	**< 0.001**
C7	3,309	63.8 (11.0)	2,025	1,284	0.889	> 0.050	0.611	**< 0.001**
C8	2,802	71.1 (10.9)	1,867	935	0.943	> 0.050	0.842	> 0.050
C9	2,581	63.7 (11.8)	1,803	778	1.314	**< 0.001**	1.789	> 0.050
C10	2,562	74.2 (9.6)	1,225	1,337	0.978	> 0.050	0.928	> 0.050
C11	2,292	66.1 (11.0)	2,186	106	0.926	> 0.050	0.650	**< 0.001**
C12	2,213	70.3 (10.2)	2,068	145	0.920	> 0.050	0.805	> 0.050
C13	2,070	58.6 (10.2)	1,348	722	0.946	> 0.050	0.577	**< 0.050**
C14	2,070	68.2 (9.6)	1,030	1,010	1.146	> 0.050	3.390	**< 0.001**
C15	2,040	63.9 (10.1)	1,208	805	1.031	> 0.050	0.784	> 0.050
C16	1,654	64.1 (12.1)	1,013	641	1.107	> 0.050	1.761	**< 0.001**
C17	1,281	65.3 (9.9)	714	567	1.001	> 0.050	1.761	**< 0.001**
C18	1,251	68.2 (9.8)	802	449	1.790	**< 0.001**	3.421	**< 0.001**
C19	1,168	58.5 (9.7)	995	173	0.752	> 0.050	1.571	> 0.050
C20	1,119	71.5 (11.3)	713	406	1.213	> 0.050	1.782	**< 0.001**
C21	1,000	61.0 (11.0)	769	231	1.116	> 0.050	0.890	> 0.050
C22	988	69.2 (10.4)	516	472	1.023	> 0.050	0.978	> 0.050
C23	935	58.7 (12.2)	588	347	1.609	**< 0.001**	2.275	**< 0.001**
C24	932	67.9 (10.1)	28	904	0.787	> 0.050	1.589	**< 0.001**
C25	860	56.2 (9.9)	664	196	0.978	> 0.050	2.691	**< 0.001**
C26	852	58.7 (12.1)	391	461	0.939	> 0.050	1.108	> 0.050
C27	823	65.1 (10.9)	532	291	1.201	> 0.050	1.289	> 0.050
C28	686	71.7 (8.0)	673	13	0.866	> 0.050	1.786	**< 0.001**
C29	550	57.2 (11.1)	435	115	0.906	> 0.050	0.985	> 0.050
C30	533	61.2 (11.7)	391	172	1.874	**< 0.001**	5.364	**< 0.001**
C31	520	64.4 (11.2)	213	307	1.052	> 0.050	1.484	> 0.050
NA*	5,113	60.1 (11.1)	3,878	1,235	NA	NA	NA	NA

*Patients that did not cluster or were in clusters of size < 500

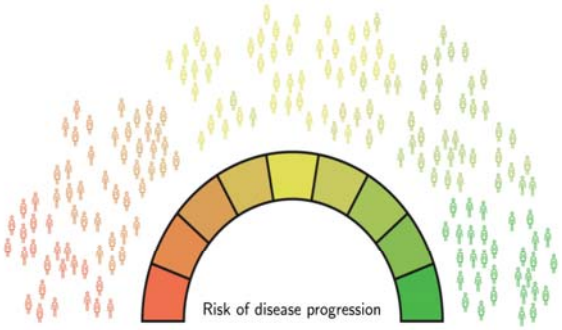
Fig 1

A. The assemblage of patient-specific vectors



B. Unsupervised clustering based on patient-specific vectors

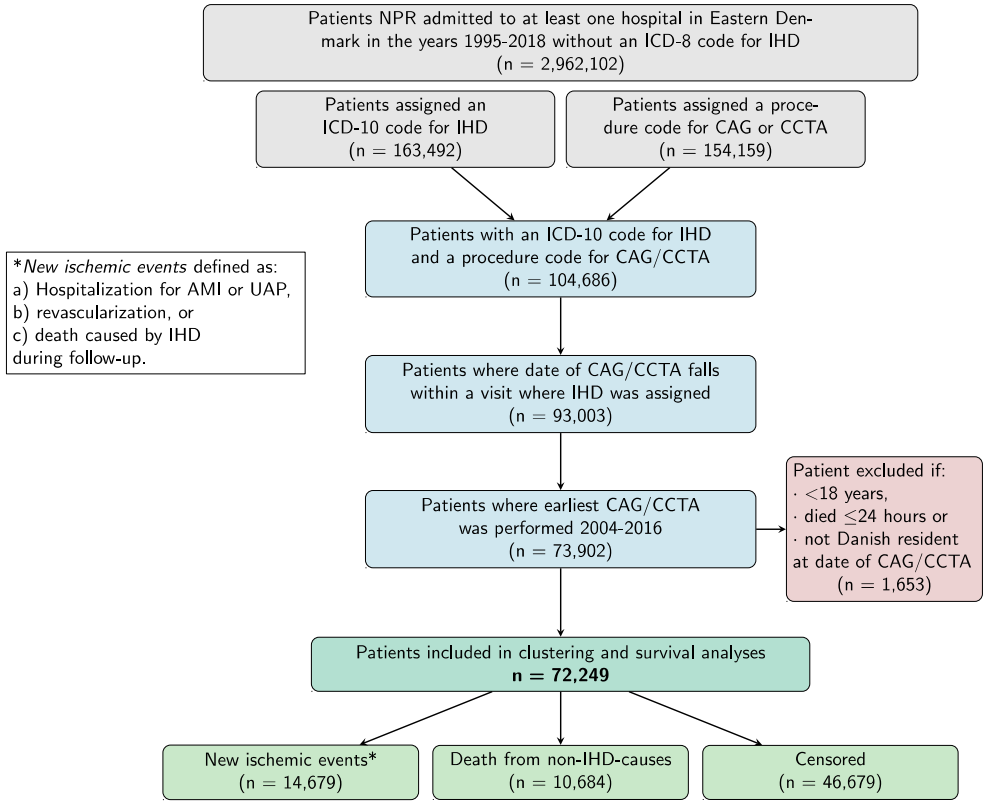
C. Association between clusters and clinical outcomes



D. Phenotypic and genetic characterization of clusters

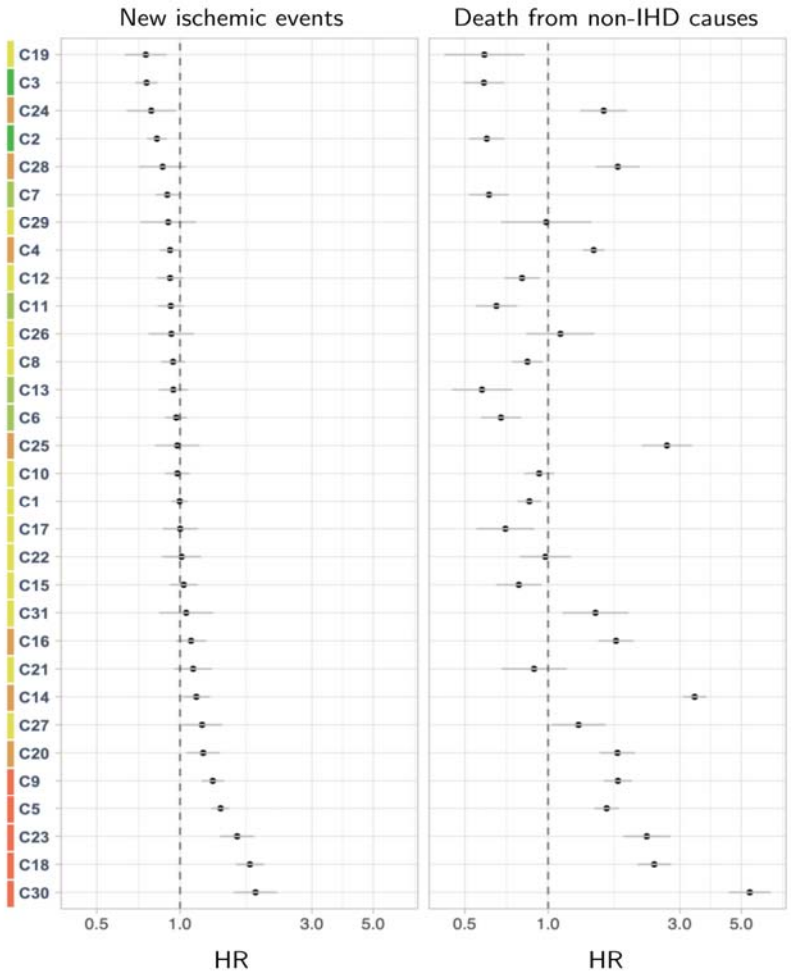
620

Fig 2


Patients NPR admitted to at least one hospital in Eastern Denmark in the years 1995-2018 without an ICD-8 code for IHD (n = 2,962,102)
Patients assigned an ICD-10 code for IHD (n = 163,492)
Patients assigned a procedure code for CAG or CCTA (n = 154,159)
Patients with an ICD-10 code for IHD and a procedure code for CAG/CCTA (n = 104,686)
*New ischemic events defined as:
a) Hospitalization for AMI or UAP,
b) revascularization, or
c) death caused by IHD
during follow-up.
Patients where date of CAG/CCTA falls within a visit where IHD was assigned (n = 93,003)
Patients where earliest CAG/CCTA was performed 2004-2016 (n = 73,902)
Patient excluded if:
· <18 years,
· died ≤24 hours or
· not Danish resident
at date of CAG/CCTA
(n = 1,653)
Patients included in clustering and survival analyses
n = 72,249
New ischemic events* (n = 14,679)
Death from non-IHD-causes (n = 10,684)
Censored (n = 46,679)

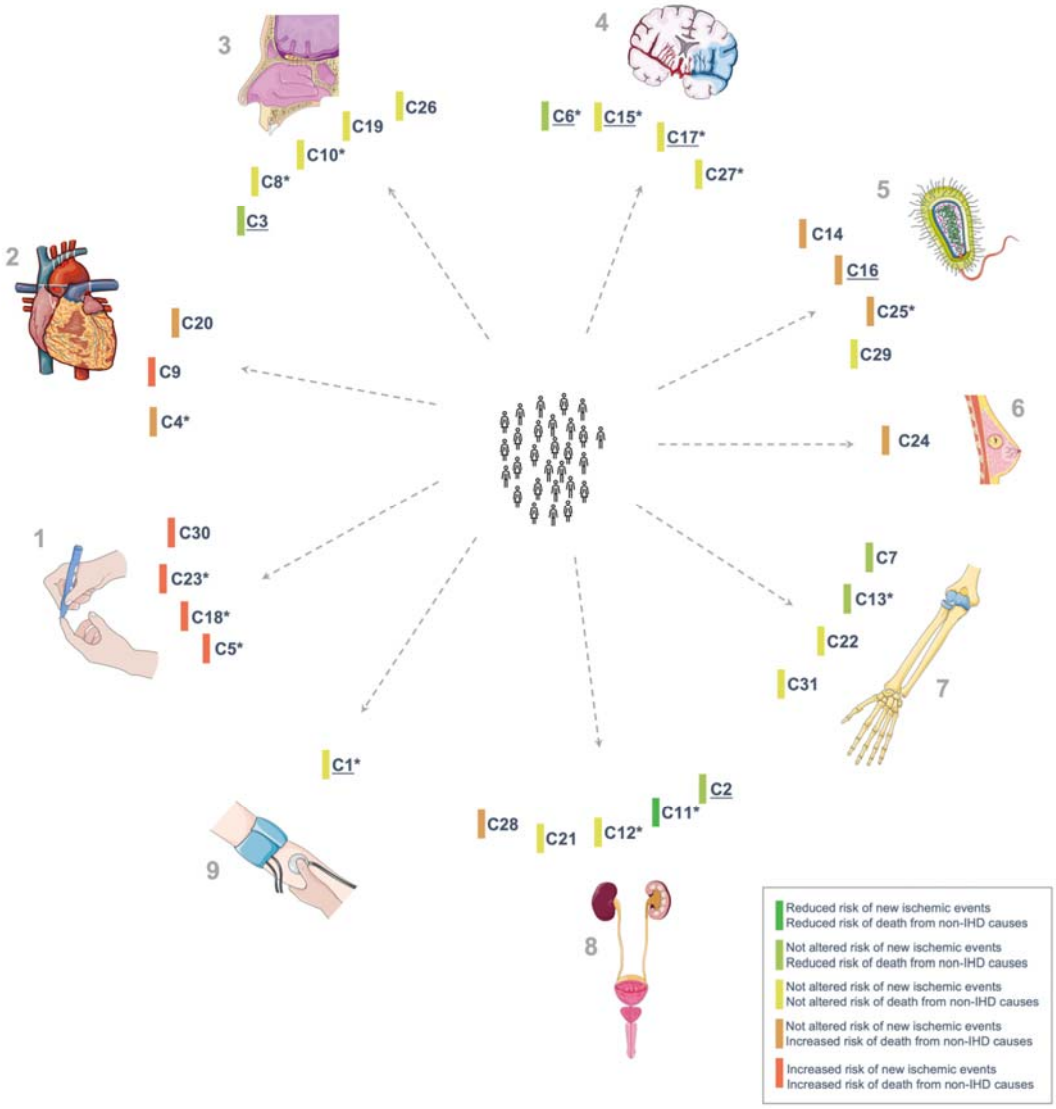
621

Fig 3


New ischemic events
Death from non-IHD causes
C19
C3
C24
C2
C28
C7
C29
C4
C12
C11
C26
C8
C13
C6
C25
C10
C1
C17
C22
C15
C31
C16
C21
C14
C27
C20
C9
C5
C23
C18
C30
0.5
1.0
3.0
5.0
HR

622

Fig 4


3
C26
C19
C10*
C8*
C3
4
C6*
C15*
C17*
C27*
5
C14
C16
C25*
C29
2
C20
C9
C4*
6
C24
1
C30
C23*
C18*
C5*
C7
C13*
C22
C31
7
C1*
C2
C11*
C28
C21
C12*
9
8
Reduced risk of new ischemic events
Reduced risk of death from non-IHD causes
Not altered risk of new ischemic events
Reduced risk of death from non-IHD causes
Not altered risk of new ischemic events
Not altered risk of death from non-IHD causes
Not altered risk of new ischemic events
Increased risk of death from non-IHD causes
Increased risk of new ischemic events
Increased risk of death from non-IHD causes

623

Supplemental material

S1 Fig: Classification of new ischemic events.

S1 Table: Eligible codes for inclusion and outcomes

S1 Appendix: Construction of patient similarity network, MCL algorithm settings and assessment of cluster robustness

- **S2 Fig: Selection of number of components.**
- **S3 Fig: Limiting edge-density and average node degree in sex-specific similarity networks**.

S2 Appendix: Preprocessing of laboratory data

- **S2 Table: Laboratory codes included in assessment of data quality and completeness**

S3 Appendix: Calculation of polygenetic risk scores for 14 traits

S4 Fig: Results of robustness analysis.

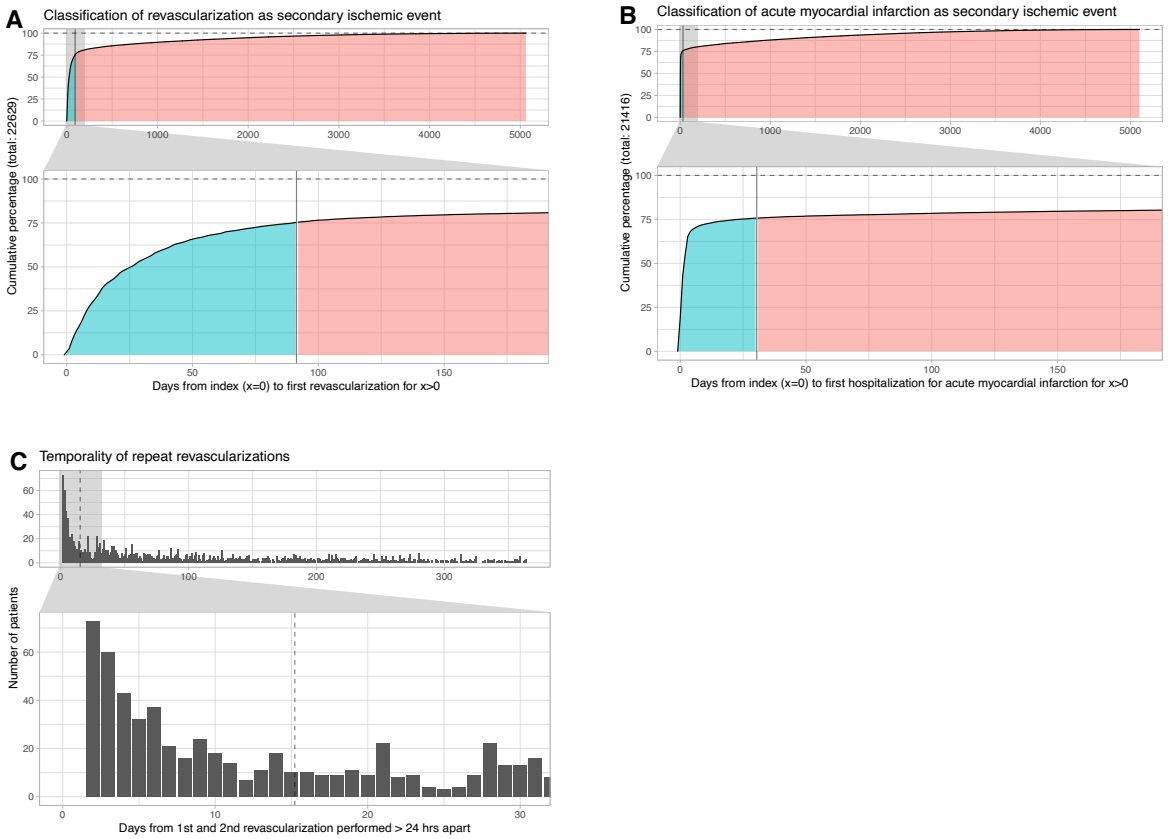
S3 Table: Comparison of mean age at index in 31 cluster using Tukey's HSD

S4 Table: Demographics for patients not cluster or were in clusters of size < 500

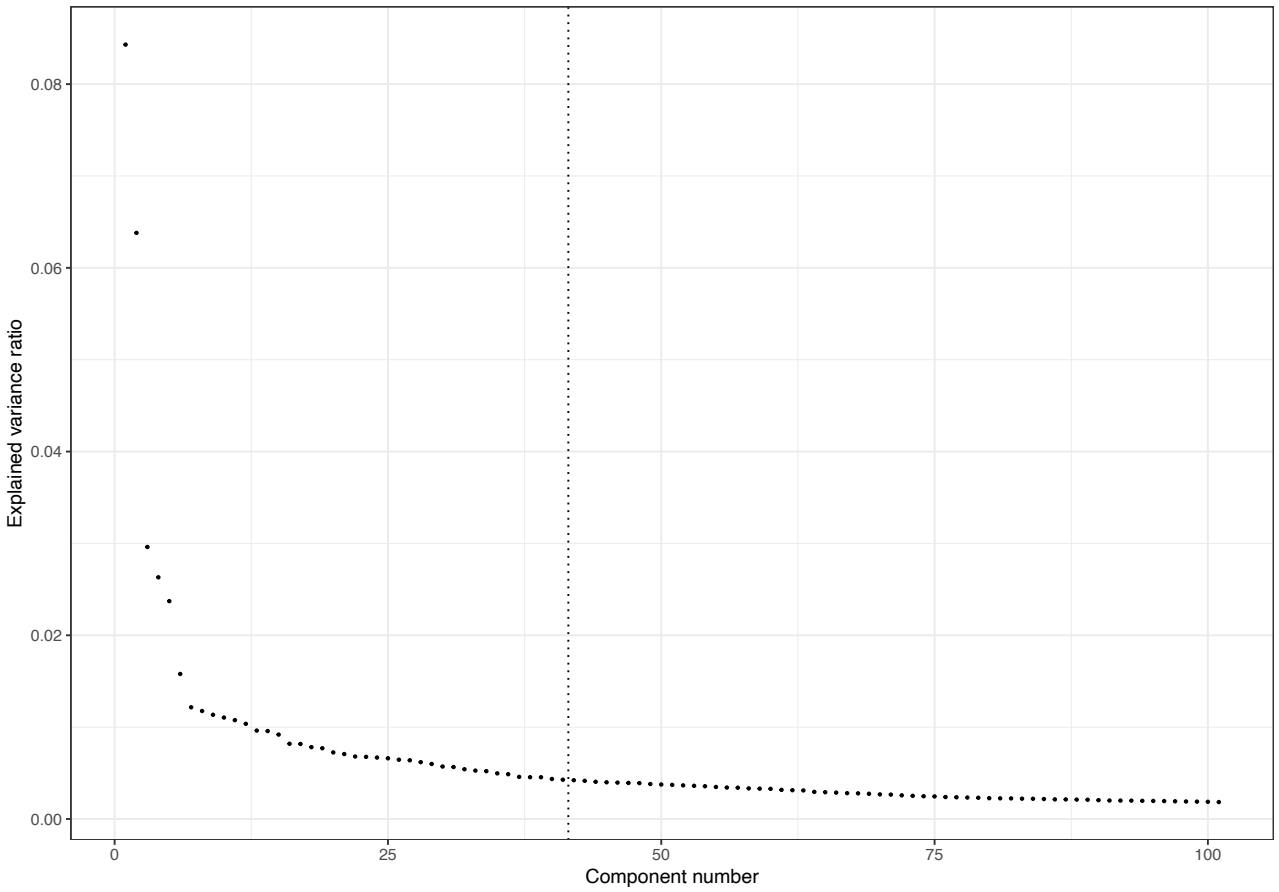
S5A-B Table: Cluster-wise summarized O/E-ratios, 10 largest O/E-ratios and 10 lowest O/E-ratios.

S6 Table: Chi-squared test for distribution laboratory values in clusters

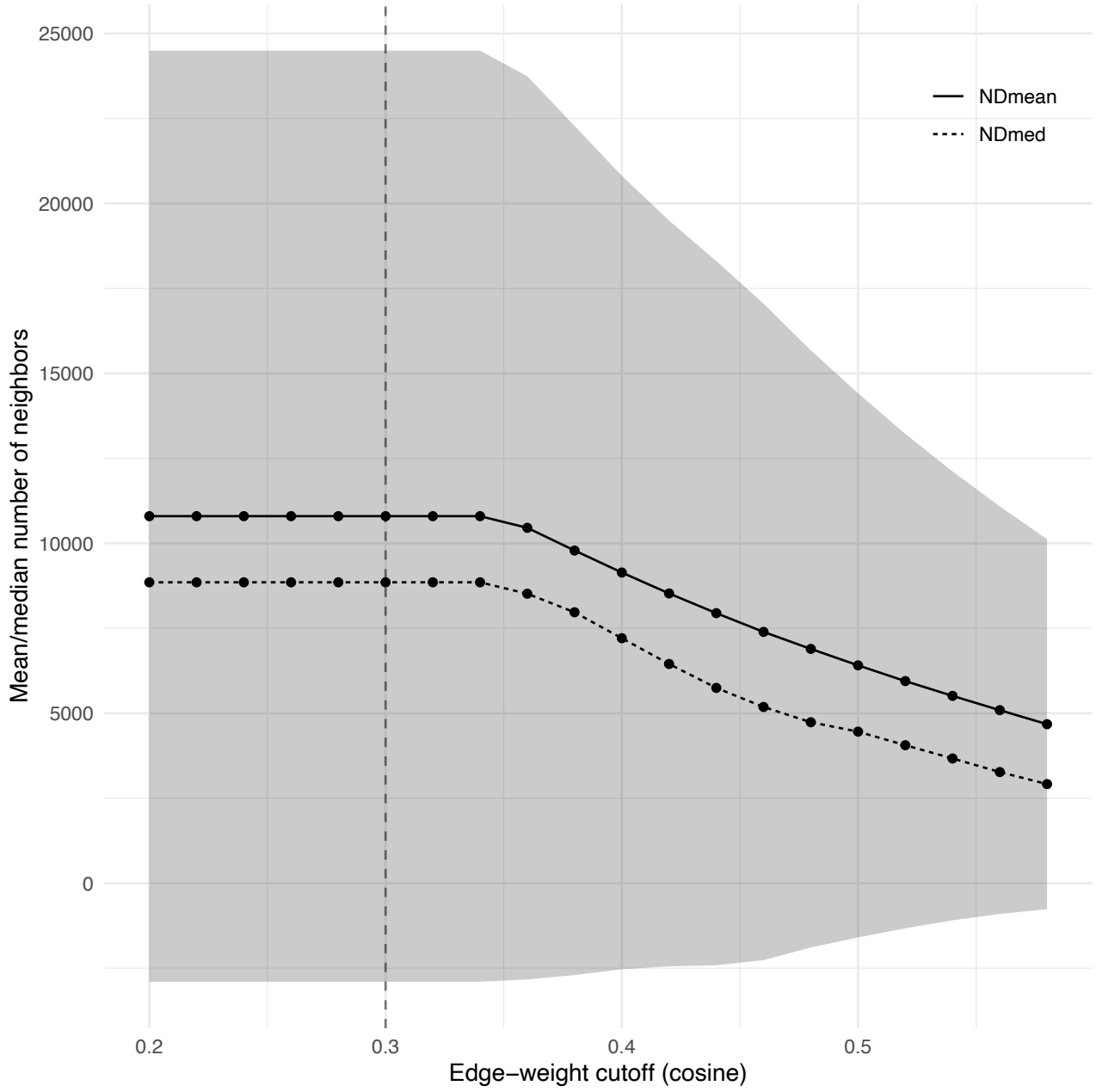
S7 Table: Traits with significantly different PGS distributions in clusters



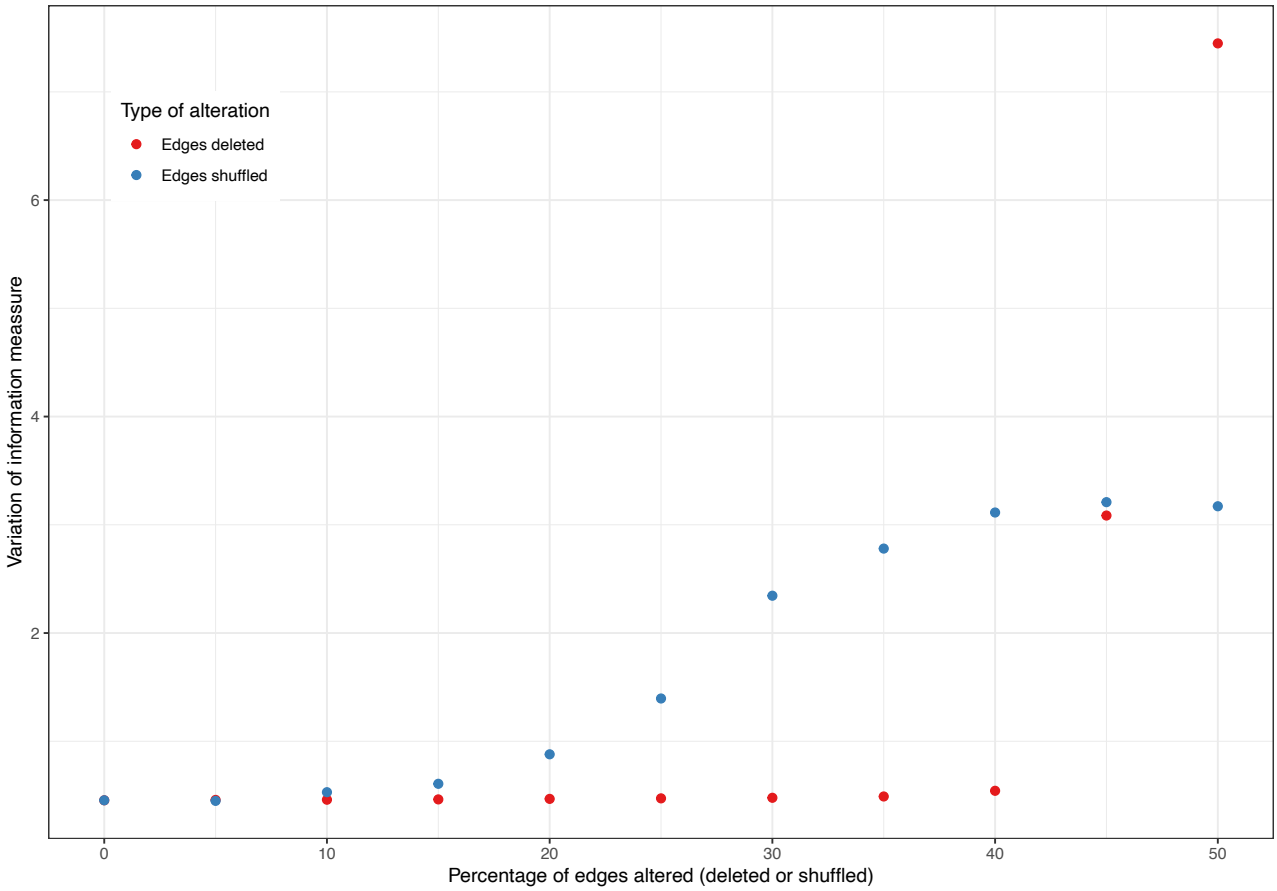
S1 Fig: Classification of new ischemic events. A: Time from index to first revascularization vs. percentage of patients revascularized. Blue corresponds to events related to establishment of IHD. Red corresponds to events considered new ischemic events. B: Time from index to first hospitalization for acute myocardial infarction vs. percentage of patients hospitalized. Blue corresponds to events related to index. Red corresponds to events considered new ischemic events. C: Distribution of days between revascularization for patients subjected to >1 performed >24 hours apart. Revascularizations performed <2 weeks apart were analyzed as a single event performed at date of the earliest revascularization. Marked by dashed line. IHD: Ischemic heart disease.



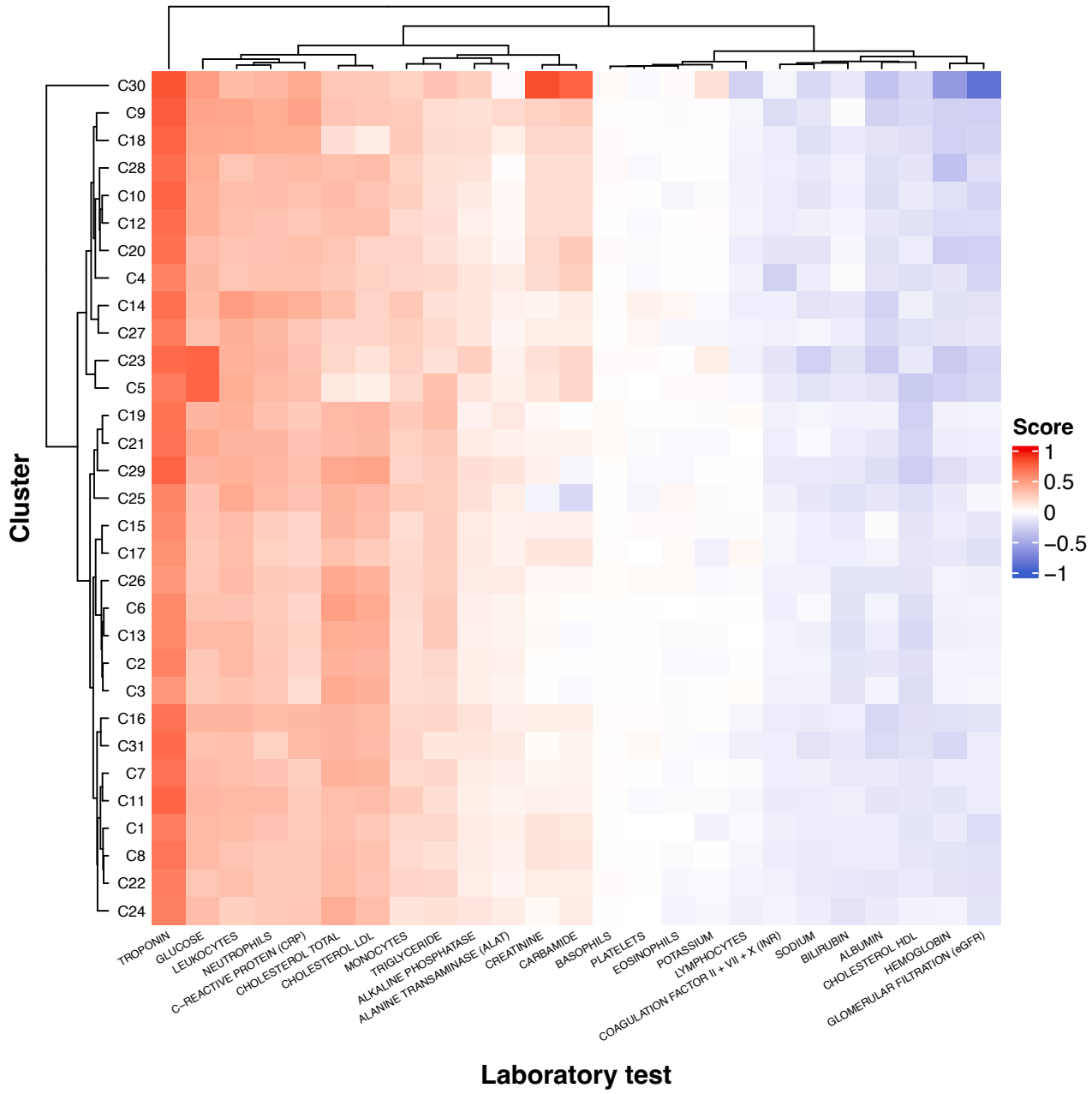
S2 Fig: Selection of number of components. X-axis: Component number ranked by explained variance ratio. Y-axis: Explained variance ratio. Dashed horizontal line indicates the cutoff.



S3 Fig: Limiting edge-density and average node degree in sex-specific similarity networks. Mean/median number of neighbors against edge-weight cutoff in the patient similarity network. Only edges with a weight higher than 0.3 (as indicated by the vertical, dashed bar) were retained.



S4 Fig: Results of robustness analysis. X-axis: Percentage of altered edges (deleted or removed). Y-axis: Variation of information measure compared to the reference graph. Legend: Type of alteration, with 10 mutations of the reference graph for each type.



S5 Fig: Heatmap of clusters based on laboratory profiles. Summary of results from the phenotypic characterization of clusters based on laboratory data. *Score* refers to the mean summarized values per cluster, where values were assigned based on the results of the laboratory test per patient. Values of -1, 0, and 1, indicates below, within or above reference range, respectively. For details, see Methods. X-axis: Laboratory test. Y-axis: Cluster.

S1 Table: Eligible codes for inclusion and outcomes

ICD-10[1] chapter IX			Definition, level 3
Block	**Level 3**	**Level 4**	
R94	I20	I20.0*, I20.1, I20.8, I20.9	Angina pectoris
	I21*	I21.0, I21.1, I21.2, I21.3, I21.4, I21.9	Acute myocardial infarction
	I23	I25.0, I25.1, I25.2, I25.23, I25.4, I25.5, I25.6, I25.8, I25.9	Certain current complications following acute myocardial infarction
	I24	I24.0, I24.1, I24.8, I24.9	Certain current complications following acute myocardial infarction
	I25	I25.0, I25.1, I25.2, I25.23, I25.4, I25.5, I25.6, I25.8, I25.9	Chronic ischemic heart disease
Nomesco[2] code		**Procedure**	
FNA*		Connection to coronary artery from internal mammary artery	
FNB*		Connection to coronary artery from gastroepiploic artery	
FNC*		Aorto-coronary venous bypass	
FND*		Aorto-coronary bypass using prosthetic graft	
FNE*		Coronary bypass using free arterial graft	
FNF*		Coronary thrombendarterectomy	
FNG*		Expansion and recanalisation of coronary artery	
SKS[3] code		**Procedure**	
UXAC85[A-D]		Coronary arteriography	
UXCC00A		Coronary computed tomography angiography	
SHAK[4] code		**Hospital**	
1301		Rigshospitalet	
1309		Bispebjerg og Frederiksberg Hospitaler	
1330		Amager og Hvidovre Hospital	
1351		Amager Hospital	
1401		Frederiksberg Hospital	
1501		Gentofte Hospital	
1502		Glostrup Hospital	
1516		Herlev og Gentofte Hospital	
2000		Hospitalerne i Nordsjælland	
2501		Amtssygehuset i Roskilde	
3800		**Region Sjællands Sygehusvæsen**	
4001		Bornholms Hospital	

[1] ICD-10 = WHO International classification of diseases and health related problems 10th edition. Danish version where code types A, B and G included in our definition of primary and secondary codes.
[2] NOMESCO = Nordic Medico-Statistical Committee
[3] SKS = Sundhedsstyrelses klassifikationsstystem [Danish]
[4] SHAK = Sygehus- og afdelingsklassifikation [Danish]
[*] Included in the composite outcome new ischemic events. For ICD-10 codes only code types A (primary) and in-hospital patients.

S2 Table: Laboratory codes included in assessment of data quality and completeness

Blood analyte	NPU codes and local systems
Sodium	NPU03429, GEN00992, NPU03796, POC00022, 240, POC00021, POC00023, GEN00990
Potassium	NPU03230, GEN00995, POC00019, POC00018, POC00020, GEN00993
Hemoglobin	NPU02319, GEN00989, NPU02321, NPU02320, NPU02322, NPU17007, POC00013, NPU04208, NPU01393, POC00012, POC00014, NPU29057, GEN00987
Creatinine / EGFR	NPU04998, NPU03918, NPU09102, NPU19661, NPU14048, NPU03800, HLL00037, DNK35131, POC00109, RHB00941, NPU28842

S3 Table: Comparison of mean age at index in 31 cluster using Tukey's HSD

contrast	null.value	estimate	conf.low	conf.high	adj.p.value
C2-C1	0	-6.18	-6.89	-5.47	0
C3-C1	0	-7.96	-8.73	-7.20	0
C4-C1	0	0,221527778	04.02	05.57	0
C5-C1	0	-0.920	-1.70	-0.139	3.60e- 3
C6-C1	0	-5.12	-5.95	-4.29	0
C7-C1	0	-0.992	-1.84	-0.141	4.40e- 3
C8-C1	0	06.29	05.39	07.20	0
C9-C1	0	-0.0638	-0.993	0,600694444	1 e+ 0
C10-C1	0	09.43	08.50	10.04	0
C11-C1	0	01.32	0,240972222	02.29	1.42e- 4
C12-C1	0	05.48	04.50	06.46	0
C13-C1	0	-6.13	-7.14	-5.12	0
C14-C1	0	03.47	02.45	04.48	0
C15-C1	0	-0.908	-1.93	0,078472222	1.79e- 1
C16-C1	0	-0.761	-1.86	0,238194444	7.29e- 1
C17-C1	0	0,355555556	-0.715	0,093055556	1.00e+ 0
C18-C1	0	03.39	02.15	0,210416667	3.19e-13
C19-C1	0	-6.27	-7.55	-4.99	0
C20-C1	0	0,301388889	05.44	08.04	0
C21-C1	0	-3.78	-5.15	-2.41	3.24e-13
C22-C1	0	04.38	03.01	0,261111111	0
C23-C1	0	-6.09	-7.50	-4.68	0
C24-C1	0	03.10	0,089583333	04.51	3.94e-13
C25-C1	0	-8.56	-10.0	-7.10	0
C26-C1	0	-5.98	-7.45	-4.52	0
C27-C1	0	0,239583333	-1.14	0,099305556	1.00e+ 0
C28-C1	0	0,313194444	05.29	08.53	0
C29-C1	0	-7.55	-9.34	-5.76	0
C30-C1	0	-3.55	-5.37	-1.73	9.14e-11
C31-C1	0	-0.335	-2.17	01.50	1.00e+ 0
C3-C2	0	-1.78	-2.58	-0.993	3.02e-13
C4-C2	0	11.00	10.02	11.08	0
C5-C2	0	05.26	04.45	06.07	0
C6-C2	0	01.06	0,140972222	0,104861111	1.32e- 3
C7-C2	0	05.19	04.31	06.06	0
C8-C2	0	12.05	11.05	13.04	0
C9-C2	0	06.11	05.16	07.07	0
C10-C2	0	15.06	14.07	16.06	0
C11-C2	0	07.50	06.50	08.49	0
C12-C2	0	11.07	10.07	12.07	0
C13-C2	0	0,356944444	-0.981	01.08	1 e+ 0
C14-C2	0	0,420138889	0,375694444	10.07	0
C15-C2	0	05.27	04.23	06.31	0
C16-C2	0	05.42	04.29	06.54	0
C17-C2	0	0,297916667	05.44	0,356944444	0
C18-C2	0	09.57	08.31	10.08	0
C19-C2	0	-0.0893	-1.38	01.21	1 e+ 0
C20-C2	0	12.09	11.06	14.02	0
C21-C2	0	02.40	01.02	0,179166667	3.07e- 8
C22-C2	0	10.06	09.17	12.00	0

contrast	null.value	estimate	conf.low	conf.high	adj.p.value
C23-C2	0	0,607638889	-1.34	01.51	1 e+ 0
C24-C2	0	09.28	0,351388889	10.07	0
C25-C2	0	-2.38	-3.86	-0.908	5.53e- 7
C26-C2	0	0,135416667	-1.29	0,088888889	1 e+ 0
C27-C2	0	06.52	05.02	08.03	0
C28-C2	0	13.01	11.05	14.07	0
C29-C2	0	-1.38	-3.18	0,297222222	5.03e- 1
C30-C2	0	0,127083333	0,555555556	04.46	2.82e- 5
C31-C2	0	0,266666667	0,19375	0,339583333	0
C4-C3	0	12.08	11.09	13.06	0
C5-C3	0	07.04	06.19	0,354166667	0
C6-C3	0	0,141666667	0,106944444	0,176388889	0
C7-C3	0	0,317361111	06.05	0,353472222	0
C8-C3	0	14.03	13.03	15.02	0
C9-C3	0	0,354166667	0,313194444	0,395138889	0
C10-C3	0	17.04	16.04	18.04	0
C11-C3	0	09.28	08.25	10.03	0
C12-C3	0	13.04	12.04	14.05	0
C13-C3	0	0,1	0,531944444	0,146527778	4.82e- 8
C14-C3	0	11.04	10.04	12.05	0
C15-C3	0	07.06	0,276388889	08.14	0
C16-C3	0	07.20	06.04	08.36	0
C17-C3	0	08.48	07.20	0,427083333	0
C18-C3	0	11.04	10.01	12.06	0
C19-C3	0	0,090277778	0,256944444	03.02	6.19e- 4
C20-C3	0	14.07	13.04	16.01	0
C21-C3	0	04.18	0,136805556	05.59	9.64e-14
C22-C3	0	12.03	10.09	13.08	0
C23-C3	0	0,102083333	0,292361111	03.32	5.07e- 4
C24-C3	0	11.01	0,417361111	12.05	0
C25-C3	0	-0.599	-2.10	0,627777778	1.00e+ 0
C26-C3	0	0,109722222	0,327083333	03.49	3.37e- 4
C27-C3	0	08.31	0,304166667	0,433333333	0
C28-C3	0	14.09	13.02	16.05	0
C29-C3	0	0,284722222	-1.42	02.24	1.00e+ 0
C30-C3	0	04.41	02.56	06.27	2.13e-13
C31-C3	0	0,335416667	0,261111111	09.50	0
C5-C4	0	-5.71	-6.58	-4.84	0
C6-C4	0	-9.91	-10.8	-9.00	0
C7-C4	0	-5.78	-6.71	-4.85	0
C8-C4	0	01.50	0,363888889	02.48	3.38e- 6
C9-C4	0	-4.85	-5.86	-3.85	0
C10-C4	0	0,211111111	0,169444444	0,253472222	0
C11-C4	0	-3.47	-4.52	-2.43	0
C12-C4	0	0,478472222	-0.366	0,093055556	8.21e- 1
C13-C4	0	-10.9	-12.0	-9.84	0
C14-C4	0	-1.32	-2.41	-0.239	1.76e- 3
C15-C4	0	-5.70	-6.79	-4.61	0
C16-C4	0	-5.55	-6.72	-4.38	0
C17-C4	0	-4.28	-5.56	-2.99	0
C18-C4	0	-1.40	-2.70	-0.102	1.70e- 2
C19-C4	0	-11.1	-12.4	-9.73	0

contrast	null.value	estimate	conf.low	conf.high	adj.p.value
C20-C4	0	0,107638889	0,411111111	03.30	2.86e- 5
C21-C4	0	-8.57	-9.99	-7.15	0
C22-C4	0	-0.409	-1.83	01.02	1.00e+ 0
C23-C4	0	-10.9	-12.3	-9.42	0
C24-C4	0	-1.69	-3.15	-0.228	5.18e- 3
C25-C4	0	-13.4	-14.9	-11.8	0
C26-C4	0	-10.8	-12.3	-9.26	0
C27-C4	0	-4.45	-5.98	-2.91	2.72e-13
C28-C4	0	02.12	0,317361111	0,179166667	6.71e- 4
C29-C4	0	-12.3	-14.2	-10.5	0
C30-C4	0	-8.34	-10.2	-6.48	0
C31-C4	0	-5.13	-7.00	-3.25	3.35e-13
C6-C5	0	-4.20	-5.12	-3.29	0
C7-C5	0	-0.0714	-1.01	0,600694444	1 e+ 0
C8-C5	0	07.21	06.23	08.20	0
C9-C5	0	0,594444444	-0.152	0,101388889	2.61e- 1
C10-C5	0	10.04	09.34	11.04	0
C11-C5	0	02.24	01.19	03.29	8.06e-13
C12-C5	0	06.40	05.34	07.46	0
C13-C5	0	-5.21	-6.29	-4.12	0
C14-C5	0	04.39	03.30	05.48	0
C15-C5	0	0,0875	-1.08	01.11	1 e+ 0
C16-C5	0	0,111111111	-1.01	01.33	1 e+ 0
C17-C5	0	01.43	0,099305556	0,133333333	1.04e- 2
C18-C5	0	04.31	03.01	0,250694444	0
C19-C5	0	-5.35	-6.68	-4.01	0
C20-C5	0	0,3375	06.30	09.02	0
C21-C5	0	-2.86	-4.28	-1.44	1.70e-11
C22-C5	0	05.30	0,185416667	0,300694444	0
C23-C5	0	-5.17	-6.63	-3.71	0
C24-C5	0	04.02	02.56	05.49	2.89e-13
C25-C5	0	-7.64	-9.16	-6.13	0
C26-C5	0	-5.06	-6.58	-3.55	0
C27-C5	0	01.27	-0.275	0,139583333	3.30e- 1
C28-C5	0	0,349305556	06.17	09.49	0
C29-C5	0	-6.63	-8.47	-4.80	0
C30-C5	0	-2.63	-4.49	-0.769	4.58e- 5
C31-C5	0	0,40625	-1.29	02.47	1.00e+ 0
C7-C6	0	04.13	03.15	05.11	0
C8-C6	0	11.04	10.04	12.04	0
C9-C6	0	05.06	04.01	06.10	0
C10-C6	0	14.06	13.05	15.06	0
C11-C6	0	06.44	05.36	07.52	0
C12-C6	0	10.06	09.51	11.07	0
C13-C6	0	-1.01	-2.12	0,077083333	1.59e- 1
C14-C6	0	08.59	07.47	0,424305556	0
C15-C6	0	04.21	03.09	05.34	0
C16-C6	0	04.36	03.16	05.56	0
C17-C6	0	0,252083333	04.32	0,315972222	0
C18-C6	0	08.51	07.18	0,433333333	0
C19-C6	0	-1.15	-2.51	0,150694444	2.81e- 1
C20-C6	0	11.09	10.05	13.02	0

contrast	null.value	estimate	conf.low	conf.high	adj.p.value
C21-C6	0	01.34	-0.107	0,138194444	1.21e- 1
C22-C6	0	09.50	08.05	11.00	0
C23-C6	0	-0.970	-2.46	0,359027778	8.23e- 1
C24-C6	0	08.22	0,301388889	0,424305556	0
C25-C6	0	-3.44	-4.98	-1.90	3.25e-13
C26-C6	0	-0.862	-2.41	0,472222222	9.65e- 1
C27-C6	0	05.47	0,1875	07.03	0
C28-C6	0	12.00	10.03	13.07	0
C29-C6	0	-2.43	-4.29	-0.579	3.36e- 4
C30-C6	0	01.57	-0.307	03.45	2.91e- 1
C31-C6	0	0,221527778	0,145138889	0,297916667	2.61e-13
C8-C7	0	07.29	06.25	08.32	0
C9-C7	0	0,644444444	-0.135	0,110416667	2.10e- 1
C10-C7	0	10.04	09.36	11.05	0
C11-C7	0	02.31	01.21	03.41	1.70e-12
C12-C7	0	06.47	05.36	07.58	0
C13-C7	0	-5.14	-6.27	-4.00	0
C14-C7	0	04.46	03.32	0,25	0
C15-C7	0	0,583333333	-1.06	01.23	1 e+ 0
C16-C7	0	0,160416667	-0.988	01.45	1.00e+ 0
C17-C7	0	01.50	0,119444444	0,141666667	7.96e- 3
C18-C7	0	04.38	03.04	0,259027778	0
C19-C7	0	-5.28	-6.65	-3.90	0
C20-C7	0	0,342361111	06.33	09.13	0
C21-C7	0	-2.79	-4.25	-1.33	3.09e-10
C22-C7	0	05.37	0,188194444	0,308333333	0
C23-C7	0	-5.10	-6.60	-3.60	0
C24-C7	0	04.10	02.59	0,25	3.44e-13
C25-C7	0	-7.57	-9.12	-6.02	0
C26-C7	0	-4.99	-6.55	-3.44	0
C27-C7	0	01.34	-0.240	0,146527778	2.64e- 1
C28-C7	0	0,354166667	06.20	0,416666667	0
C29-C7	0	-6.56	-8.43	-4.70	0
C30-C7	0	-2.56	-4.45	-0.668	1.53e- 4
C31-C7	0	0,45625	-1.25	02.57	1.00e+ 0
C9-C8	0	-6.36	-7.46	-5.25	0
C10-C8	0	03.14	02.03	04.25	3.03e-13
C11-C8	0	-4.98	-6.12	-3.84	0
C12-C8	0	-0.813	-1.96	0,234722222	6.78e- 1
C13-C8	0	-12.4	-13.6	-11.2	0
C14-C8	0	-2.83	-4.00	-1.65	2.02e-13
C15-C8	0	-7.20	-8.38	-6.02	0
C16-C8	0	-7.05	-8.31	-5.80	0
C17-C8	0	-5.78	-7.15	-4.42	0
C18-C8	0	-2.90	-4.28	-1.52	1.41e-12
C19-C8	0	-12.6	-14.0	-11.2	0
C20-C8	0	0,309027778	-0.987	0,102777778	1.00e+ 0
C21-C8	0	-10.1	-11.6	-8.58	0
C22-C8	0	-1.91	-3.41	-0.414	6.53e- 4
C23-C8	0	-12.4	-13.9	-10.9	0
C24-C8	0	-3.19	-4.72	-1.66	2.46e-12
C25-C8	0	-14.9	-16.4	-13.3	0

contrast	null.value	estimate	conf.low	conf.high	adj.p.value
C26-C8	0	-12.3	-13.9	-10.7	0
C27-C8	0	-5.95	-7.55	-4.34	0
C28-C8	0	0,427777778	-1.11	02.34	1.00e+ 0
C29-C8	0	-13.8	-15.7	-12.0	0
C30-C8	0	-9.84	-11.8	-7.93	0
C31-C8	0	-6.63	-8.56	-4.70	0
C10-C9	0	09.50	08.37	10.06	0
C11-C9	0	01.38	0,152777778	02.54	3.01e- 3
C12-C9	0	05.54	04.37	0,3	0
C13-C9	0	-6.06	-7.26	-4.87	0
C14-C9	0	03.53	02.33	0,217361111	1.45e-13
C15-C9	0	-0.844	-2.05	0,25	6.94e- 1
C16-C9	0	-0.697	-1.97	0,401388889	9.74e- 1
C17-C9	0	0,4	-0.808	0,108333333	1.00e+ 0
C18-C9	0	03.46	02.06	0,225694444	3.13e-13
C19-C9	0	-6.20	-7.63	-4.78	0
C20-C9	0	0,305555556	05.35	08.25	0
C21-C9	0	-3.72	-5.22	-2.21	3.27e-13
C22-C9	0	04.45	0,147916667	0,275	1.72e-13
C23-C9	0	-6.03	-7.57	-4.48	0
C24-C9	0	03.17	0,084722222	0,215972222	6.14e-12
C25-C9	0	-8.50	-10.1	-6.91	0
C26-C9	0	-5.92	-7.52	-4.32	0
C27-C9	0	0,284027778	-1.21	02.03	1.00e+ 0
C28-C9	0	0,317361111	05.23	0,382638889	0
C29-C9	0	-7.49	-9.39	-5.59	0
C30-C9	0	-3.48	-5.41	-1.56	4.48e- 9
C31-C9	0	-0.271	-2.22	0,088888889	1 e+ 0
C11-C10	0	-8.12	-9.28	-6.95	0
C12-C10	0	-3.95	-5.13	-2.78	0
C13-C10	0	-15.6	-16.8	-14.4	0
C14-C10	0	-5.97	-7.17	-4.76	0
C15-C10	0	-10.3	-11.5	-9.13	0
C16-C10	0	-10.2	-11.5	-8.92	0
C17-C10	0	-8.92	-10.3	-7.54	0
C18-C10	0	-6.04	-7.44	-4.64	0
C19-C10	0	-15.7	-17.1	-14.3	0
C20-C10	0	-2.69	-4.15	-1.24	1.25e- 9
C21-C10	0	-13.2	-14.7	-11.7	0
C22-C10	0	-5.05	-6.57	-3.54	0
C23-C10	0	-15.5	-17.1	-14.0	0
C24-C10	0	-6.33	-7.88	-4.78	0
C25-C10	0	-18.0	-19.6	-16.4	0
C26-C10	0	-15.4	-17.0	-13.8	0
C27-C10	0	-9.09	-10.7	-7.47	0
C28-C10	0	-2.52	-4.26	-0.783	2.17e- 5
C29-C10	0	-17.0	-18.9	-15.1	0
C30-C10	0	-13.0	-14.9	-11.1	0
C31-C10	0	-9.77	-11.7	-7.82	0
C12-C11	0	04.16	0,15	05.37	0
C13-C11	0	-7.45	-8.67	-6.22	0
C14-C11	0	02.15	0,636805556	03.38	2.38e- 8

contrast	null.value	estimate	conf.low	conf.high	adj.p.value
C15-C11	0	-2.23	-3.46	-0.989	5.77e- 9
C16-C11	0	-2.08	-3.38	-0.772	9.60e- 7
C17-C11	0	-0.806	-2.22	0,421527778	9.55e- 1
C18-C11	0	02.07	0,452083333	03.50	1.84e- 5
C19-C11	0	-7.59	-9.04	-6.13	0
C20-C11	0	05.42	0,190277778	0,3125	0
C21-C11	0	-5.10	-6.63	-3.56	0
C22-C11	0	03.06	01.52	0,208333333	3.32e-11
C23-C11	0	-7.41	-8.98	-5.84	0
C24-C11	0	0,096527778	0,147916667	03.36	7.19e- 3
C25-C11	0	-9.88	-11.5	-8.26	0
C26-C11	0	-7.30	-8.93	-5.68	0
C27-C11	0	-0.973	-2.62	0,466666667	9.33e- 1
C28-C11	0	05.59	0,182638889	07.35	0
C29-C11	0	-8.87	-10.8	-6.95	0
C30-C11	0	-4.87	-6.81	-2.92	2.88e-13
C31-C11	0	-1.65	-3.62	0,218055556	2.82e- 1
C13-C12	0	-11.6	-12.8	-10.4	0
C14-C12	0	-2.01	-3.26	-0.770	4.92e- 7
C15-C12	0	-6.39	-7.63	-5.14	0
C16-C12	0	-6.24	-7.56	-4.93	0
C17-C12	0	-4.97	-6.39	-3.55	0
C18-C12	0	-2.09	-3.52	-0.656	1.82e- 5
C19-C12	0	-11.7	-13.2	-10.3	0
C20-C12	0	01.26	-0.227	0,134722222	2.66e- 1
C21-C12	0	-9.26	-10.8	-7.72	0
C22-C12	0	-1.10	-2.65	0,3125	6.69e- 1
C23-C12	0	-11.6	-13.2	-9.99	0
C24-C12	0	-2.38	-3.96	-0.796	6.89e- 6
C25-C12	0	-14.0	-15.7	-12.4	0
C26-C12	0	-11.5	-13.1	-9.83	0
C27-C12	0	-5.14	-6.79	-3.48	0
C28-C12	0	01.43	-0.340	03.20	3.67e- 1
C29-C12	0	-13.0	-15.0	-11.1	0
C30-C12	0	-9.03	-11.0	-7.08	0
C31-C12	0	-5.82	-7.79	-3.84	1.31e-13
C14-C13	0	09.59	08.33	10.09	0
C15-C13	0	05.22	0,190972222	06.49	0
C16-C13	0	05.37	04.03	0,298611111	0
C17-C13	0	0,294444444	05.20	08.08	0
C18-C13	0	09.52	08.07	11.00	0
C19-C13	0	-0.141	-1.62	01.34	1 e+ 0
C20-C13	0	12.09	11.04	14.04	0
C21-C13	0	02.35	0,547222222	0,188194444	6.61e- 6
C22-C13	0	10.05	0,398611111	12.01	0
C23-C13	0	0,250694444	-1.56	0,085416667	1 e+ 0
C24-C13	0	09.23	0,335416667	10.08	0
C25-C13	0	-2.44	-4.08	-0.793	1.07e- 5
C26-C13	0	0,1	-1.50	0,096527778	1 e+ 0
C27-C13	0	06.47	0,222222222	08.14	0
C28-C13	0	13.00	11.03	14.08	0
C29-C13	0	-1.43	-3.37	0,357638889	5.91e- 1

contrast	null.value	estimate	conf.low	conf.high	adj.p.value
C30-C13	0	02.58	0,425	04.54	3.42e- 4
C31-C13	0	0,263194444	0,18125	0,345833333	2.55e-13
C15-C14	0	-4.37	-5.65	-3.10	0
C16-C14	0	-4.23	-5.57	-2.89	0
C17-C14	0	-2.96	-4.40	-1.51	6.09e-12
C18-C14	0	-0.0752	-1.53	01.38	1 e+ 0
C19-C14	0	-9.74	-11.2	-8.25	0
C20-C14	0	03.27	0,094444444	0,220833333	5.00e-13
C21-C14	0	-7.25	-8.81	-5.68	0
C22-C14	0	0,634722222	-0.655	02.48	9.43e- 1
C23-C14	0	-9.56	-11.2	-7.96	0
C24-C14	0	-0.364	-1.96	01.24	1.00e+ 0
C25-C14	0	-12.0	-13.7	-10.4	0
C26-C14	0	-9.45	-11.1	-7.80	0
C27-C14	0	-3.12	-4.79	-1.45	9.54e-10
C28-C14	0	03.44	0,0875	05.23	1.89e-10
C29-C14	0	-11.0	-13.0	-9.08	0
C30-C14	0	-7.02	-8.99	-5.05	0
C31-C14	0	-3.80	-5.79	-1.81	2.86e-10
C16-C15	0	0,102083333	-1.20	01.49	1 e+ 0
C17-C15	0	01.42	-0.0271	0,14375	6.33e- 2
C18-C15	0	04.30	0,141666667	0,261111111	1.24e-13
C19-C15	0	-5.36	-6.85	-3.87	0
C20-C15	0	0,336805556	06.14	09.16	0
C21-C15	0	-2.87	-4.44	-1.31	2.33e- 9
C22-C15	0	05.29	0,175	0,309722222	0
C23-C15	0	-5.18	-6.79	-3.58	0
C24-C15	0	04.01	02.41	0,251388889	2.86e-13
C25-C15	0	-7.66	-9.30	-6.01	0
C26-C15	0	-5.08	-6.73	-3.42	0
C27-C15	0	01.25	-0.422	0,147916667	5.49e- 1
C28-C15	0	0,348611111	06.03	0,417361111	0
C29-C15	0	-6.65	-8.59	-4.70	0
C30-C15	0	-2.64	-4.61	-0.669	2.01e- 4
C31-C15	0	0,397916667	-1.42	02.56	1.00e+ 0
C17-C16	0	01.27	-0.234	0,1375	2.71e- 1
C18-C16	0	04.15	0,127777778	0,254861111	3.14e-13
C19-C16	0	-5.51	-7.05	-3.96	0
C20-C16	0	07.50	0,272916667	09.07	0
C21-C16	0	-3.02	-4.64	-1.40	1.10e- 9
C22-C16	0	05.14	03.51	0,303472222	0
C23-C16	0	-5.33	-6.99	-3.67	0
C24-C16	0	0,184722222	02.21	05.52	2.42e-13
C25-C16	0	-7.80	-9.50	-6.10	0
C26-C16	0	-5.22	-6.93	-3.52	0
C27-C16	0	01.11	-0.621	0,140972222	8.50e- 1
C28-C16	0	0,338194444	0,265972222	09.51	0
C29-C16	0	-6.79	-8.79	-4.80	0
C30-C16	0	-2.79	-4.80	-0.772	8.57e- 5
C31-C16	0	0,295833333	-1.61	02.46	1.00e+ 0
C18-C17	0	0,144444444	01.27	04.49	7.47e- 9
C19-C17	0	-6.78	-8.42	-5.14	0

contrast	null.value	estimate	conf.low	conf.high	adj.p.value
C20-C17	0	06.23	04.57	0,352777778	0
C21-C17	0	-4.29	-6.00	-2.58	2.71e-13
C22-C17	0	0,185416667	02.16	05.58	2.99e-13
C23-C17	0	-6.60	-8.34	-4.86	0
C24-C17	0	02.59	0,588888889	04.33	9.90e- 6
C25-C17	0	-9.08	-10.9	-7.29	0
C26-C17	0	-6.50	-8.29	-4.71	0
C27-C17	0	-0.167	-1.98	0,086111111	1 e+ 0
C28-C17	0	06.40	04.48	08.31	0
C29-C17	0	-8.07	-10.1	-6.00	0
C30-C17	0	-4.06	-6.15	-1.97	1.10e-10
C31-C17	0	-0.847	-2.95	01.26	1.00e+ 0
C19-C18	0	-9.66	-11.3	-8.01	0
C20-C18	0	03.35	0,088888889	05.01	1.86e-11
C21-C18	0	-7.17	-8.89	-5.46	0
C22-C18	0	0,686805556	-0.734	0,132638889	9.52e- 1
C23-C18	0	-9.48	-11.2	-7.73	0
C24-C18	0	-0.289	-2.04	01.46	1 e+ 0
C25-C18	0	-12.0	-13.7	-10.2	0
C26-C18	0	-9.38	-11.2	-7.58	0
C27-C18	0	-3.05	-4.86	-1.23	1.26e- 7
C28-C18	0	03.52	01.59	05.44	2.68e- 9
C29-C18	0	-10.9	-13.0	-8.87	0
C30-C18	0	-6.94	-9.03	-4.85	0
C31-C18	0	-3.73	-5.84	-1.61	1.43e- 8
C20-C19	0	13.00	11.03	14.07	0
C21-C19	0	02.49	0,516666667	04.23	3.54e- 5
C22-C19	0	10.06	0,395833333	12.04	0
C23-C19	0	0,122916667	-1.60	0,107638889	1 e+ 0
C24-C19	0	09.37	07.59	11.01	0
C25-C19	0	-2.29	-4.11	-0.476	8.54e- 4
C26-C19	0	0,197222222	-1.54	02.11	1 e+ 0
C27-C19	0	0,292361111	0,220138889	08.46	0
C28-C19	0	13.02	11.02	15.01	0
C29-C19	0	-1.29	-3.38	0,561111111	8.99e- 1
C30-C19	0	0,133333333	0,41875	0,225	5.58e- 4
C31-C19	0	0,272916667	0,180555556	08.07	2.91e-13
C21-C20	0	-10.5	-12.3	-8.76	0
C22-C20	0	-2.36	-4.12	-0.589	2.25e- 4
C23-C20	0	-12.8	-14.6	-11.0	0
C24-C20	0	-3.63	-5.43	-1.84	1.18e-11
C25-C20	0	-15.3	-17.1	-13.5	0
C26-C20	0	-12.7	-14.6	-10.9	0
C27-C20	0	-6.39	-8.25	-4.53	0
C28-C20	0	0,11875	-1.79	02.13	1 e+ 0
C29-C20	0	-14.3	-16.4	-12.2	0
C30-C20	0	-10.3	-12.4	-8.16	0
C31-C20	0	-7.07	-9.22	-4.92	0
C22-C21	0	08.16	06.35	0,443055556	0
C23-C21	0	-2.31	-4.15	-0.469	9.66e- 4
C24-C21	0	0,311111111	05.04	0,384027778	0
C25-C21	0	-4.78	-6.67	-2.90	2.43e-13

contrast	null.value	estimate	conf.low	conf.high	adj.p.value
C26-C21	0	-2.20	-4.09	-0.316	4.30e- 3
C27-C21	0	04.13	02.22	06.03	5.42e-13
C28-C21	0	10.07	0,380555556	12.07	0
C29-C21	0	-3.77	-5.92	-1.62	1.78e- 8
C30-C21	0	0,161111111	-1.94	02.40	1 e+ 0
C31-C21	0	03.45	01.26	0,252083333	1.42e- 6
C23-C22	0	-10.5	-12.3	-8.63	0
C24-C22	0	-1.28	-3.13	0,396527778	7.22e- 1
C25-C22	0	-12.9	-14.8	-11.1	0
C26-C22	0	-10.4	-12.3	-8.47	0
C27-C22	0	-4.04	-5.95	-2.13	1.26e-12
C28-C22	0	02.53	0,358333333	04.54	9.36e- 4
C29-C22	0	-11.9	-14.1	-9.78	0
C30-C22	0	-7.93	-10.1	-5.75	0
C31-C22	0	-4.72	-6.91	-2.52	6.49e-13
C24-C23	0	09.19	07.32	11.01	0
C25-C23	0	-2.47	-4.38	-0.559	4.88e- 4
C26-C23	0	0,074305556	-1.81	02.02	1 e+ 0
C27-C23	0	06.44	04.50	08.37	0
C28-C23	0	13.00	11.00	15.00	0
C29-C23	0	-1.46	-3.64	0,495138889	7.74e- 1
C30-C23	0	02.54	0,239583333	0,218055556	5.09e- 3
C31-C23	0	0,261111111	03.54	0,359027778	3.28e-13
C25-C24	0	-11.7	-13.6	-9.75	0
C26-C24	0	-9.09	-11.0	-7.17	0
C27-C24	0	-2.76	-4.69	-0.822	3.68e- 5
C28-C24	0	0,18125	0,095138889	0,266666667	9.20e-10
C29-C24	0	-10.7	-12.8	-8.48	0
C30-C24	0	-6.65	-8.85	-4.45	0
C31-C24	0	-3.44	-5.65	-1.22	2.37e- 6
C26-C25	0	02.58	0,431944444	04.54	3.02e- 4
C27-C25	0	0,396527778	0,314583333	10.09	0
C28-C25	0	15.05	13.04	17.05	0
C29-C25	0	01.01	-1.20	03.22	9.98e- 1
C30-C25	0	05.01	0,1375	07.25	3.14e-13
C31-C25	0	08.23	0,276388889	10.05	0
C27-C26	0	06.33	04.35	08.31	0
C28-C26	0	12.09	10.08	15.00	0
C29-C26	0	-1.57	-3.78	0,447222222	6.70e- 1
C30-C26	0	02.44	0,138194444	0,213194444	1.45e- 2
C31-C26	0	0,253472222	03.40	0,354166667	2.79e-13
C28-C27	0	06.56	04.47	0,379166667	0
C29-C27	0	-7.90	-10.1	-5.67	0
C30-C27	0	-3.89	-6.14	-1.64	3.41e- 8
C31-C27	0	-0.680	-2.95	01.59	1.00e+ 0
C29-C28	0	-14.5	-16.8	-12.1	0
C30-C28	0	-10.5	-12.8	-8.12	0
C31-C28	0	-7.24	-9.60	-4.89	0
C30-C29	0	04.01	01.54	06.47	4.11e- 7
C31-C29	0	07.22	0,218055556	0,423611111	2.62e-13
C31-C30	0	03.21	0,499305556	0,257638889	5.28e- 4
M-F	0	-2.75	-2.91	-2.58	0

S4 Table: Demographics for patients that did not cluster or were in clusters of size < 500

Cohort demographics	**Total**	**Males**	**Females**
Number of patients	5,113	3,878	1,235
Mean age at index (SD)	60.7	60.0	63.0
Outcomes, number of cases	**Total**	**Males**	**Females**
New ischemic events	995	780	175
Death from non-IHD causes	352	274	78
Censored	3,624	2,707	917
Outcomes, time to event	**Mean time to event in years (SD)**		
	Total	**Males**	**Females**
New ischemic events	1.55 (1.41)	1.59 (1.43)	1.39 (1.32)
Death from non-IHD causes	2.25 (1.50)	2.18 (1.47)	2.5 (1.49)
Censored	4.54 (0.95)	4.52 (0.96)	2.47 (1.49)
Total	4.02 (1.52)	3.98 (1.54)	4.17 (1.44)

S5A Table: Degree of enrichment (sum) and top-10 O/E-ratios per cluster

Clst	Sum	ICD-10	# code	# clst	Obs	Exp	O/E-ratio	description
C1	345	I109	24818	7191	1.000	0.294	3.40	Essential (primary) hypertension
C1	345	H350	548	7191	0.016	0.007	2.24	Background retinopathy and retinal vascular changes
C1	345	E871	393	7191	0.012	0.005	2.23	Hypo-osmolality and hyponatraemia
C1	345	I159	414	7191	0.012	0.005	2.15	Secondary hypertension, unspecified
C1	345	I959	283	7191	0.008	0.004	2.15	Hypotension, unspecified
C1	345	E789	417	7191	0.012	0.006	2.10	Disorder of lipoprotein metabolism, unspecified
C1	345	E785	5002	7191	0.133	0.067	1.98	Hyperlipidaemia, unspecified
C1	345	I119	534	7191	0.013	0.007	1.85	Hypertensive heart disease without (congestive) heart failure
C1	345	D251	256	7191	0.006	0.004	1.83	Intramural leiomyoma of uterus
C1	345	N811	974	7191	0.024	0.013	1.77	Cystocele
C2	372	K802	2849	5990	0.176	0.029	6.01	Calculus of gallbladder without cholecystitis
C2	372	K801	411	5990	0.023	0.004	5.22	Calculus of gallbladder with other cholecystitis
C2	372	K805	1027	5990	0.054	0.011	4.70	Calculus of bile duct without cholangitis or cholecystitis
C2	372	K800	468	5990	0.022	0.006	3.88	Calculus of gallbladder with acute cholecystitis
C2	372	R100	2884	5990	0.123	0.035	3.51	Acute abdomen
C2	372	R108	3700	5990	0.158	0.045	3.51	Abdominal and pelvic pain
C2	372	R103	488	5990	0.020	0.006	3.29	Pain localized to other parts of lower abdomen
C2	372	R102	566	5990	0.023	0.007	3.29	Pelvic and perineal pain
C2	372	N832	489	5990	0.020	0.006	3.28	Other and unspecified ovarian cysts
C2	372	D251	256	5990	0.010	0.003	3.19	Intramural leiomyoma of uterus
C3	268	R079	5863	4641	0.363	0.067	5.43	Pain in throat and chest
C3	268	G409	841	4641	0.043	0.010	4.23	Epilepsy, unspecified
C3	268	I309	297	4641	0.014	0.004	3.70	Acute pericarditis, unspecified
C3	268	M626	4440	4641	0.187	0.057	3.28	Muscle strain
C3	268	G430	351	4641	0.013	0.005	2.83	Migraine without aura [common migraine]
C3	268	R073	2009	4641	0.072	0.027	2.67	Other chest pain
C3	268	R002	691	4641	0.025	0.009	2.66	Palpitations
C3	268	R519	1667	4641	0.058	0.022	2.61	Headache
C3	268	R064	542	4641	0.019	0.007	2.54	Hyperventilation
C3	268	R072	367	4641	0.012	0.005	2.48	Precordial pain
C4	520	I489	7075	4401	0.995	0.043	23.14	Atrial fibrillation and atrial flutter, unspecified
C4	520	I495	482	4401	0.055	0.004	14.14	Sick sinus syndrome
C4	520	I480	364	4401	0.039	0.003	12.49	Paroxysmal atrial fibrillation
C4	520	I471	1920	4401	0.183	0.018	10.27	Supraventricular tachycardia
C4	520	I499	436	4401	0.035	0.005	7.71	Cardiac arrhythmia, unspecified
C4	520	I479	614	4401	0.045	0.007	6.73	Paroxysmal tachycardia, unspecified
C4	520	R001	394	4401	0.026	0.004	5.80	Bradycardia, unspecified
C4	520	R000	391	4401	0.025	0.004	5.51	Tachycardia, unspecified
C4	520	I340	971	4401	0.051	0.012	4.32	Mitral (valve) insufficiency
C4	520	I491	273	4401	0.013	0.003	3.76	Atrial premature depolarization
C5	596	E119	7551	4290	0.947	0.056	17.06	Type 2 diabetes mellitus: Without complications
C5	596	E113	720	4290	0.081	0.006	13.55	Type 2 diabetes mellitus: With ophthalmic complications
C5	596	E114	881	4290	0.088	0.008	11.06	Type 2 diabetes mellitus: With neurological complications
C5	596	E149	958	4290	0.095	0.009	10.82	Unspecified diabetes mellitus: Without complications

Clst	Sum	ICD-10	# code	# clst	Obs	Exp	O/E-ratio	description
C5	596	E117	980	4290	0.096	0.009	10.58	Type 2 diabetes mellitus: With multiple complications
C5	596	E118	2669	4290	0.251	0.025	9.93	Type 2 diabetes mellitus: With unspecified complications
C5	596	H360	1467	4290	0.137	0.014	9.80	Diabetic retinopathy
C5	596	E112	761	4290	0.070	0.007	9.53	Type 2 diabetes mellitus: With renal complications
C5	596	E148	483	4290	0.044	0.005	9.50	Unspecified diabetes mellitus: With unspecified complications
C5	596	E115	551	4290	0.048	0.005	8.88	Type 2 diabetes mellitus: With peripheral circulatory complications
C6	239	E780	12780	3589	0.867	0.152	5.70	Pure hypercholesterolaemia
C6	239	E785	5002	3589	0.177	0.069	2.57	Hyperlipidaemia, unspecified
C6	239	I999	383	3589	0.010	0.005	1.89	Other and unspecified disorders of circulatory system
C6	239	I639	1989	3589	0.044	0.029	1.53	Cerebral infarction, unspecified
C6	239	I652	563	3589	0.011	0.008	1.39	Occlusion and stenosis of carotid artery
C6	239	G459	2066	3589	0.042	0.030	1.39	Transient cerebral ischaemic attack, unspecified
C6	239	R670	750	3589	0.014	0.011	1.24	NA
C6	239	E113	720	3589	0.013	0.011	1.21	Type 2 diabetes mellitus: With ophthalmic complications
C6	239	M100	691	3589	0.012	0.010	1.20	Idiopathic gout
C6	239	N434	294	3589	0.005	0.004	1.16	Spermatocele
C7	374	M171	2940	3309	0.359	0.027	13.06	Other primary gonarthrosis
C7	374	M179	2242	3309	0.257	0.022	11.80	Gonarthrosis, unspecified
C7	374	M170	2145	3309	0.232	0.022	10.74	Primary gonarthrosis, bilateral
C7	374	M234	258	3309	0.027	0.003	10.33	Loose body in knee
C7	374	M235	269	3309	0.028	0.003	10.19	Chronic instability of knee
C7	374	M232	2404	3309	0.238	0.025	9.42	Derangement of meniscus due to old tear or injury
C7	374	M238	532	3309	0.042	0.006	6.89	Other internal derangements of knee
C7	374	M239	1105	3309	0.081	0.013	6.21	Internal derangement of knee, unspecified
C7	374	M712	363	3309	0.025	0.004	5.63	Synovial cyst of popliteal space [Baker]
C7	374	M169	1539	3309	0.088	0.020	4.48	Coxarthrosis, unspecified
C8	384	H919	4610	2802	0.664	0.043	15.54	Hearing loss, unspecified
C8	384	H911	3527	2802	0.495	0.033	14.90	Presbycusis
C8	384	H905	1160	2802	0.151	0.011	13.13	Sensorineural hearing loss, unspecified
C8	384	H833	1412	2802	0.180	0.014	12.70	Noise effects on inner ear
C8	384	H838	254	2802	0.032	0.003	12.38	Other specified diseases of inner ear
C8	384	H938	1116	2802	0.132	0.012	11.34	Other specified disorders of ear
C8	384	H908	631	2802	0.071	0.007	10.50	Mixed conductive and sensorineural hearing loss, unspecified
C8	384	H931	1228	2802	0.123	0.014	9.01	Tinnitus
C8	384	H810	300	2802	0.029	0.003	8.64	Ménière disease
C8	384	H809	374	2802	0.034	0.004	7.82	Otosclerosis, unspecified
C9	323	I420	706	2581	0.095	0.007	13.21	Dilated cardiomyopathy
C9	323	I509	6160	2581	0.783	0.064	12.21	Heart failure, unspecified
C9	323	I429	479	2581	0.056	0.005	10.75	Cardiomyopathy, unspecified
C9	323	I501	1502	2581	0.124	0.018	6.74	Left ventricular failure
C9	323	I500	2327	2581	0.188	0.029	6.57	Congestive heart failure
C9	323	R570	320	2581	0.025	0.004	6.38	Cardiogenic shock
C9	323	I460	1028	2581	0.075	0.013	5.78	Cardiac arrest with successful resuscitation
C9	323	I472	752	2581	0.053	0.010	5.52	Ventricular tachycardia

Clst	Sum	ICD-10	# code	# clst	Obs	Exp	O/E-ratio	description
C9	323	I110	273	2581	0.018	0.004	5.20	Hypertensive heart disease with (congestive) heart failure
C9	323	I340	971	2581	0.059	0.013	4.68	Mitral (valve) insufficiency
C10	361	H259	5764	2562	0.866	0.055	15.77	Senile cataract, unspecified
C10	361	H264	1015	2562	0.113	0.011	10.03	After-cataract
C10	361	H330	418	2562	0.045	0.005	9.68	Retinal detachment with retinal break
C10	361	H353	1743	2562	0.174	0.020	8.67	Degeneration of macula and posterior pole
C10	361	H401	388	2562	0.032	0.005	6.86	Primary open-angle glaucoma
C10	361	H438	291	2562	0.022	0.004	6.14	Other disorders of vitreous body
C10	361	H333	264	2562	0.019	0.003	5.74	Retinal breaks without detachment
C10	361	H521	366	2562	0.023	0.005	4.84	Myopia
C10	361	H348	337	2562	0.017	0.005	3.79	Other retinal vascular occlusions
C10	361	H260	486	2562	0.023	0.007	3.55	Infantile, juvenile and presenile cataract
C11	328	K409	3787	2292	0.984	0.024	41.64	Unilateral or unspecified inguinal hernia, without obstruction or gangrene
C11	328	K402	280	2292	0.042	0.003	14.76	Bilateral inguinal hernia, without obstruction or gangrene
C11	328	N433	367	2292	0.023	0.005	4.67	Hydrocele, unspecified
C11	328	N434	294	2292	0.010	0.004	2.40	Spermatocele
C11	328	K429	896	2292	0.030	0.013	2.32	Umbilical hernia without obstruction or gangrene
C11	328	N484	456	2292	0.011	0.007	1.71	Impotence of organic origin
C11	328	N508	300	2292	0.007	0.004	1.70	Other specified disorders of male genital organs
C11	328	M720	1004	2292	0.024	0.015	1.67	Palmar fascial fibromatosis [Dupuytren]
C11	328	D179	257	2292	0.006	0.004	1.63	Benign lipomatous neoplasm, unspecified
C11	328	I714	517	2292	0.012	0.008	1.56	Abdominal aortic aneurysm, without mention of rupture
C12	383	N409	3319	2213	0.701	0.027	25.77	Hyperplasia of prostate
C12	383	R339	1530	2213	0.239	0.015	15.55	Retention of urine
C12	383	R391	2230	2213	0.241	0.026	9.24	Other difficulties with micturition
C12	383	N359	253	2213	0.027	0.003	9.12	Urethral stricture, unspecified
C12	383	R319	3787	2213	0.323	0.047	6.82	Unspecified haematuria
C12	383	N459	612	2213	0.052	0.008	6.72	Orchitis, epididymitis and epididymo-orchitis without abscess
C12	383	M720	1004	2213	0.059	0.013	4.36	Palmar fascial fibromatosis [Dupuytren]
C12	383	N434	294	2213	0.015	0.004	3.71	Spermatocele
C12	383	N309	834	2213	0.042	0.011	3.64	Cystitis, unspecified
C12	383	N319	265	2213	0.013	0.004	3.60	Neuromuscular dysfunction of bladder, unspecified
C13	522	M511	3357	2070	0.937	0.022	42.98	Lumbar and other intervertebral disc disorders with radiculopathy
C13	522	M519	604	2070	0.120	0.005	22.05	Intervertebral disc disorder, unspecified
C13	522	M512	366	2070	0.071	0.003	21.34	Other specified intervertebral disc displacement
C13	522	M544	1209	2070	0.144	0.014	10.28	Lumbago with sciatica
C13	522	M543	491	2070	0.057	0.006	9.94	Sciatica
C13	522	M513	1585	2070	0.178	0.019	9.54	Other specified intervertebral disc degeneration
C13	522	M539	274	2070	0.026	0.003	7.54	Dorsopathy, unspecified
C13	522	M472	897	2070	0.082	0.011	7.35	Other spondylosis with radiculopathy
C13	522	M501	910	2070	0.080	0.011	7.01	Cervical disc disorder with radiculopathy
C13	522	M549	1419	2070	0.103	0.019	5.58	Dorsalgia, unspecified
C14	676	J440	743	2040	0.191	0.005	35.25	Chronic obstructive pulmonary disease with acute lower respiratory infection

Clst	Sum	ICD-10	# code	# clst	Obs	Exp	O/E-ratio	description
C14	676	J441	1678	2040	0.423	0.013	33.71	Chronic obstructive pulmonary disease with acute exacerbation, unspecified
C14	676	J449	4621	2040	0.948	0.041	22.97	Chronic obstructive pulmonary disease, unspecified
C14	676	J439	317	2040	0.064	0.003	22.18	Emphysema, unspecified
C14	676	J448	396	2040	0.075	0.004	20.31	Other specified chronic obstructive pulmonary disease
C14	676	J429	714	2040	0.105	0.008	13.66	Unspecified chronic bronchitis
C14	676	J960	837	2040	0.120	0.009	13.13	Acute respiratory failure
C14	676	J969	681	2040	0.072	0.008	8.78	Respiratory failure, unspecified
C14	676	J159	1222	2040	0.103	0.016	6.62	Bacterial pneumonia, unspecified
C14	676	J209	527	2040	0.041	0.007	6.05	Acute bronchitis, unspecified
C15	296	E780	12780	2013	1.000	0.165	6.05	Pure hypercholesterolaemia
C15	296	I109	24818	2013	1.000	0.350	2.86	Essential (primary) hypertension
C15	296	D629	333	2013	0.009	0.005	1.96	Acute posthaemorrhagic anaemia
C15	296	I639	1989	2013	0.055	0.029	1.91	Cerebral infarction, unspecified
C15	296	I693	467	2013	0.013	0.007	1.91	Sequelae of cerebral infarction
C15	296	R670	750	2013	0.020	0.011	1.87	NA
C15	296	E114	881	2013	0.021	0.013	1.66	Type 2 diabetes mellitus: With neurological complications
C15	296	G459	2066	2013	0.049	0.030	1.61	Transient cerebral ischaemic attack, unspecified
C15	296	E118	2669	2013	0.063	0.039	1.60	Type 2 diabetes mellitus: With unspecified complications
C15	296	R072	367	2013	0.008	0.005	1.57	Precordial pain
C16	365	J189	5496	1654	0.743	0.065	11.40	Pneumonia, unspecified
C16	365	J849	265	1654	0.027	0.003	8.10	Interstitial pulmonary disease, unspecified
C16	365	C349	284	1654	0.026	0.004	7.06	Malignant neoplasm: Bronchus or lung, unspecified
C16	365	R919	1471	1654	0.103	0.020	5.17	Abnormal findings on diagnostic imaging of lung
C16	365	J909	291	1654	0.019	0.004	4.72	Pleural effusion, not elsewhere classified
C16	365	J181	298	1654	0.018	0.004	4.43	Lobar pneumonia, unspecified
C16	365	J159	1222	1654	0.073	0.017	4.35	Bacterial pneumonia, unspecified
C16	365	R091	418	1654	0.023	0.006	3.96	Pleurisy
C16	365	J180	257	1654	0.013	0.004	3.71	Bronchopneumonia, unspecified
C16	365	J969	681	1654	0.031	0.010	3.27	Respiratory failure, unspecified
C17	375	E780	12780	1281	1.000	0.175	5.73	Pure hypercholesterolaemia
C17	375	E789	417	1281	0.018	0.006	3.00	Disorder of lipoprotein metabolism, unspecified
C17	375	H350	548	1281	0.023	0.008	2.98	Background retinopathy and retinal vascular changes
C17	375	G459	2066	1281	0.087	0.030	2.95	Transient cerebral ischaemic attack, unspecified
C17	375	I639	1989	1281	0.084	0.029	2.92	Cerebral infarction, unspecified
C17	375	I109	24818	1281	1.000	0.357	2.80	Essential (primary) hypertension
C17	375	I652	563	1281	0.022	0.008	2.69	Occlusion and stenosis of carotid artery
C17	375	I693	467	1281	0.017	0.007	2.54	Sequelae of cerebral infarction
C17	375	M109	547	1281	0.019	0.008	2.36	Gout, unspecified
C17	375	I694	1733	1281	0.059	0.025	2.36	Sequelae of stroke, not specified as haemorrhage or infarction
C18	535	I702	2251	1251	0.812	0.019	43.33	Atherosclerosis of arteries of extremities
C18	535	I739	2027	1251	0.544	0.020	26.65	Peripheral vascular disease, unspecified
C18	535	I709	433	1251	0.081	0.005	16.02	Generalized and unspecified atherosclerosis
C18	535	L979	605	1251	0.091	0.007	12.23	Ulcer of lower limb, not elsewhere classified
C18	535	E105	416	1251	0.060	0.005	11.58	Type 1 diabetes mellitus: With peripheral circulatory complications
C18	535	E115	551	1251	0.066	0.007	9.34	Type 2 diabetes mellitus: With peripheral circulatory complications

Clst	Sum	ICD-10	# code	# clst	Obs	Exp	O/E-ratio	description
C18	535	E148	483	1251	0.042	0.007	6.35	Unspecified diabetes mellitus: With unspecified complications
C18	535	I714	517	1251	0.043	0.007	6.14	Abdominal aortic aneurysm, without mention of rupture
C18	535	L984	299	1251	0.021	0.004	5.02	Chronic ulcer of skin, not elsewhere classified
C18	535	I999	383	1251	0.023	0.005	4.31	Other and unspecified disorders of circulatory system
C19	374	G473	1897	1168	0.712	0.016	44.12	Sleep apnoea
C19	374	R065	983	1168	0.271	0.010	26.88	Mouth breathing
C19	374	J342	1097	1168	0.218	0.013	17.10	Deviated nasal septum
C19	374	J330	351	1168	0.038	0.005	8.10	Polyp of nasal cavity
C19	374	J320	513	1168	0.033	0.007	4.65	Chronic maxillary sinusitis
C19	374	J350	315	1168	0.015	0.005	3.22	Chronic tonsillitis
C19	374	N508	300	1168	0.014	0.004	3.18	Other specified disorders of male genital organs
C19	374	E669	2114	1168	0.086	0.031	2.80	Obesity, unspecified
C19	374	M766	339	1168	0.013	0.005	2.62	Achilles tendinitis
C19	374	G510	365	1168	0.013	0.005	2.42	Bell's palsy
C20	337	I350	2664	1119	0.856	0.026	33.13	Aortic (valve) stenosis
C20	337	I351	696	1119	0.090	0.009	10.02	Aortic (valve) insufficiency
C20	337	K053	360	1119	0.023	0.005	4.59	Chronic periodontitis
C20	337	K045	333	1119	0.021	0.005	4.58	Chronic apical periodontitis
C20	337	R040	1768	1119	0.100	0.025	3.99	Epistaxis
C20	337	D649	1637	1119	0.084	0.023	3.59	Anaemia, unspecified
C20	337	I340	971	1119	0.045	0.014	3.20	Mitral (valve) insufficiency
C20	337	M353	627	1119	0.028	0.009	3.07	Polymyalgia rheumatica
C20	337	D509	459	1119	0.020	0.007	2.97	Iron deficiency anaemia, unspecified
C20	337	K921	287	1119	0.012	0.004	2.80	Melaena
C21	532	N200	1391	1000	0.765	0.009	80.82	Calculus of kidney
C21	532	N201	1381	1000	0.605	0.012	51.56	Calculus of ureter
C21	532	N209	520	1000	0.202	0.005	42.01	Urinary calculus, unspecified
C21	532	N133	267	1000	0.045	0.003	13.41	Other and unspecified hydronephrosis
C21	532	N109	555	1000	0.047	0.008	6.12	Acute tubulo-interstitial nephritis
C21	532	R319	3787	1000	0.197	0.054	3.63	Unspecified haematuria
C21	532	N359	253	1000	0.009	0.004	2.44	Urethral stricture, unspecified
C21	532	N308	384	1000	0.013	0.006	2.32	Other cystitis
C21	532	R100	2884	1000	0.096	0.042	2.28	Acute abdomen
C21	532	A419	968	1000	0.032	0.014	2.26	Sepsis, unspecified
C22	592	M480	2424	988	0.932	0.023	41.03	Spinal stenosis
C22	592	M431	583	988	0.135	0.007	19.79	Spondylolisthesis
C22	592	M472	897	988	0.159	0.011	14.20	Other spondylosis with radiculopathy
C22	592	M513	1585	988	0.200	0.021	9.56	Other specified intervertebral disc degeneration
C22	592	M539	274	988	0.031	0.004	8.54	Dorsopathy, unspecified
C22	592	M478	630	988	0.072	0.008	8.50	Other spondylosis
C22	592	M479	531	988	0.059	0.007	8.21	Spondylosis, unspecified
C22	592	M543	491	988	0.046	0.007	6.76	Sciatica
C22	592	M503	520	988	0.039	0.007	5.43	Other cervical disc degeneration
C22	592	R522	1177	988	0.088	0.016	5.34	Other chronic pain
C23	968	E103	576	935	0.339	0.004	86.66	Type 1 diabetes mellitus: With ophthalmic complications
C23	968	E107	583	935	0.303	0.005	66.79	Type 1 diabetes mellitus: With multiple complications

Clst	Sum	ICD-10	# code	# clst	Obs	Exp	O/E-ratio	description
C23	968	E104	434	935	0.214	0.004	60.52	Type 1 diabetes mellitus: With neurological complications
C23	968	E102	377	935	0.165	0.003	48.90	Type 1 diabetes mellitus: With renal complications
C23	968	E162	510	935	0.199	0.005	40.65	Hypoglycaemia, unspecified
C23	968	E108	1073	935	0.363	0.011	32.70	Type 1 diabetes mellitus: With unspecified complications
C23	968	E109	2680	935	0.887	0.028	31.71	Type 1 diabetes mellitus: Without complications
C23	968	E105	416	935	0.137	0.004	31.47	Type 1 diabetes mellitus: With peripheral circulatory complications
C23	968	H360	1467	935	0.398	0.017	24.05	Diabetic retinopathy
C23	968	H431	334	935	0.061	0.004	14.57	Vitreous haemorrhage
C24	559	C509	1103	932	0.835	0.005	170.04	Malignant neoplasm: Breast, unspecified
C24	559	N639	773	932	0.153	0.010	16.12	Unspecified lump in breast
C24	559	D249	758	932	0.097	0.010	9.57	Benign neoplasm of breast
C24	559	N602	712	932	0.076	0.010	7.87	Fibroadenosis of breast
C24	559	N950	699	932	0.059	0.010	6.07	Postmenopausal bleeding
C24	559	L905	295	932	0.020	0.004	4.89	Scar conditions and fibrosis of skin
C24	559	M819	1292	932	0.083	0.018	4.50	Osteoporosis, unspecified
C24	559	N629	571	932	0.036	0.008	4.50	Hypertrophy of breast
C24	559	N840	546	932	0.030	0.008	3.84	Polyp of corpus uteri
C24	559	E052	322	932	0.014	0.005	2.99	Thyrotoxicosis with toxic multinodular goitre
C25	681	F103	518	860	0.331	0.004	94.26	Mental and behavioural disorders due to use of alcohol: Withdrawal state
C25	681	F102	1189	860	0.641	0.010	66.56	Mental and behavioural disorders due to use of alcohol: Dependence syndrome
C25	681	F100	1212	860	0.535	0.011	47.14	Mental and behavioural disorders due to use of alcohol: Acute intoxication
C25	681	F101	1115	860	0.430	0.011	38.27	Mental and behavioural disorders due to use of alcohol: Harmful use
C25	681	F172	349	860	0.090	0.004	21.82	Mental and behavioural disorders due to use of tobacco: Dependence syndrome
C25	681	F339	299	860	0.053	0.004	14.01	Recurrent depressive disorder, unspecified
C25	681	R568	477	860	0.067	0.006	10.67	Other and unspecified convulsions
C25	681	K920	428	860	0.057	0.006	9.96	Haematemesis
C25	681	F329	816	860	0.093	0.011	8.38	Depressive episode, unspecified
C25	681	F419	335	860	0.036	0.005	7.86	Anxiety disorder, unspecified
C26	540	J459	2487	852	0.927	0.026	36.22	Asthma, unspecified
C26	540	J451	417	852	0.133	0.005	28.92	Nonallergic asthma
C26	540	J450	401	852	0.127	0.004	28.68	Predominantly allergic asthma
C26	540	J448	396	852	0.040	0.005	7.31	Other specified chronic obstructive pulmonary disease
C26	540	J209	527	852	0.046	0.007	6.22	Acute bronchitis, unspecified
C26	540	J330	351	852	0.029	0.005	5.97	Polyp of nasal cavity
C26	540	J370	343	852	0.028	0.005	5.85	Chronic laryngitis
C26	540	R490	415	852	0.023	0.006	3.94	Dysphonia
C26	540	J429	714	852	0.039	0.010	3.77	Unspecified chronic bronchitis
C26	540	J441	1678	852	0.083	0.024	3.44	Chronic obstructive pulmonary disease with acute exacerbation, unspecified
C27	440	I649	2991	823	0.666	0.037	18.07	Stroke, not specified as haemorrhage or infarction
C27	440	I694	1733	823	0.360	0.022	16.60	Sequelae of stroke, not specified as haemorrhage or infarction
C27	440	I693	467	823	0.079	0.006	13.03	Sequelae of cerebral infarction

Clst	Sum	ICD-10	# code	# clst	Obs	Exp	O/E-ratio	description
C27	440	I639	1989	823	0.332	0.026	12.82	Cerebral infarction, unspecified
C27	440	I652	563	823	0.085	0.007	11.44	Occlusion and stenosis of carotid artery
C27	440	G459	2066	823	0.270	0.028	9.70	Transient cerebral ischaemic attack, unspecified
C27	440	R670	750	823	0.077	0.010	7.39	NA
C27	440	G409	841	823	0.080	0.012	6.86	Epilepsy, unspecified
C27	440	G969	710	823	0.055	0.010	5.45	Disorder of central nervous system, unspecified
C27	440	R298	1366	823	0.104	0.019	5.41	Other and unspecified symptoms and signs involving the nervous and musculoskeletal systems
C28	483	C619	1173	686	0.943	0.008	119.15	Malignant neoplasm of prostate
C28	483	R339	1530	686	0.131	0.022	6.05	Retention of urine
C28	483	N484	456	686	0.031	0.007	4.68	Impotence of organic origin
C28	483	C443	522	686	0.026	0.008	3.46	Malignant neoplasm: Skin of other and unspecified parts of face
C28	483	N133	267	686	0.013	0.004	3.38	Other and unspecified hydronephrosis
C28	483	N479	462	686	0.022	0.007	3.25	Redundant prepuce, phimosis and paraphimosis
C28	483	N409	3319	686	0.147	0.048	3.04	Hyperplasia of prostate
C28	483	R391	2230	686	0.096	0.033	2.95	Other difficulties with micturition
C28	483	A419	968	686	0.039	0.014	2.78	Sepsis, unspecified
C28	483	N359	253	686	0.010	0.004	2.76	Urethral stricture, unspecified
C29	264	A630	320	550	0.082	0.004	19.81	Anogenital (venereal) warts
C29	264	L022	506	550	0.076	0.007	10.96	Cutaneous abscess, furuncle and carbuncle of trunk
C29	264	L024	953	550	0.135	0.013	10.19	Cutaneous abscess, furuncle and carbuncle of limb
C29	264	L089	1001	550	0.113	0.014	7.99	Local infection of skin and subcutaneous tissue, unspecified
C29	264	L029	622	550	0.069	0.009	7.88	Cutaneous abscess, furuncle and carbuncle, unspecified
C29	264	I803	1509	550	0.142	0.021	6.60	Phlebitis and thrombophlebitis of lower extremities, unspecified
C29	264	A499	492	550	0.044	0.007	6.21	Bacterial infection, unspecified
C29	264	A469	1352	550	0.109	0.019	5.62	Erysipelas
C29	264	I829	297	550	0.016	0.004	3.78	Embolism and thrombosis of unspecified vein
C29	264	R509	907	550	0.047	0.013	3.57	Fever, unspecified
C30	1434	N180	454	533	0.610	0.002	314.82	Chronic kidney disease
C30	1434	N199	711	533	0.585	0.006	97.71	Unspecified kidney failure
C30	1434	N189	1094	533	0.848	0.010	87.98	Chronic kidney disease, unspecified
C30	1434	K650	264	533	0.148	0.003	53.36	Acute peritonitis
C30	1434	E102	377	533	0.135	0.005	29.50	Type 1 diabetes mellitus: With renal complications
C30	1434	N179	377	533	0.086	0.005	17.37	Acute renal failure, unspecified
C30	1434	E112	761	533	0.167	0.010	16.55	Type 2 diabetes mellitus: With renal complications
C30	1434	K053	360	533	0.077	0.005	16.06	Chronic periodontitis
C30	1434	E107	583	533	0.120	0.008	15.41	Type 1 diabetes mellitus: With multiple complications
C30	1434	N133	267	533	0.045	0.004	12.34	Other and unspecified hydronephrosis
C31	727	M059	682	520	0.679	0.005	137.45	Seropositive rheumatoid arthritis, unspecified
C31	727	M069	662	520	0.567	0.006	102.97	Rheumatoid arthritis, unspecified
C31	727	M060	398	520	0.246	0.004	60.73	Seronegative rheumatoid arthritis
C31	727	M204	350	520	0.063	0.005	13.34	Other hammer toe(s) (acquired)
C31	727	M139	488	520	0.067	0.007	9.90	Arthritis, unspecified
C31	727	M029	304	520	0.031	0.004	7.12	Reactive arthropathy, unspecified
C31	727	J849	265	520	0.019	0.004	5.02	Interstitial pulmonary disease, unspecified
C31	727	M201	858	520	0.060	0.012	4.80	Hallux valgus (acquired)

Clst	Sum	ICD-10	# code	# clst	Obs	Exp	O/E-ratio	description
C31	727	M255	842	520	0.050	0.012	4.08	Pain in joint
C31	727	M190	837	520	0.048	0.012	3.94	Primary arthrosis of other joints

S5B, Table: Bottom-10 O/E-ratios < 1 per cluster

Clst	ICD-10	# code	# clst	Obs	Exp	O/E-ratio	description
C1	E107	583	7191	0.001	0.010	0.12	Type 1 diabetes mellitus: With multiple complications
C1	D303	600	7191	0.002	0.010	0.20	Benign neoplasm: Bladder
C1	E105	416	7191	0.001	0.007	0.20	Type 1 diabetes mellitus: With peripheral circulatory complications
C1	J441	1678	7191	0.006	0.027	0.21	Chronic obstructive pulmonary disease with acute exacerbation, unspecified
C1	E780	12780	7191	0.046	0.208	0.22	Pure hypercholesterolaemia
C1	E103	576	7191	0.002	0.009	0.22	Type 1 diabetes mellitus: With ophthalmic complications
C1	C619	1173	7191	0.004	0.019	0.23	Malignant neoplasm of prostate
C1	E102	377	7191	0.001	0.006	0.23	Type 1 diabetes mellitus: With renal complications
C1	H360	1467	7191	0.006	0.024	0.23	Diabetic retinopathy
C1	E104	434	7191	0.002	0.007	0.24	Type 1 diabetes mellitus: With neurological complications
C2	D303	600	5990	0.000	0.010	0.02	Benign neoplasm: Bladder
C2	C619	1173	5990	0.000	0.019	0.02	Malignant neoplasm of prostate
C2	E107	583	5990	0.000	0.010	0.02	Type 1 diabetes mellitus: With multiple complications
C2	E103	576	5990	0.000	0.009	0.02	Type 1 diabetes mellitus: With ophthalmic complications
C2	N180	454	5990	0.000	0.007	0.02	Chronic kidney disease
C2	H360	1467	5990	0.001	0.024	0.03	Diabetic retinopathy
C2	E108	1073	5990	0.001	0.017	0.03	Type 1 diabetes mellitus: With unspecified complications
C2	N409	3319	5990	0.002	0.054	0.03	Hyperplasia of prostate
C2	E104	434	5990	0.000	0.007	0.05	Type 1 diabetes mellitus: With neurological complications
C2	E148	483	5990	0.001	0.008	0.06	Unspecified diabetes mellitus: With unspecified complications
C3	E112	761	4641	0.000	0.012	0.02	Type 2 diabetes mellitus: With renal complications
C3	E113	720	4641	0.000	0.012	0.02	Type 2 diabetes mellitus: With ophthalmic complications
C3	D303	600	4641	0.000	0.010	0.02	Benign neoplasm: Bladder
C3	C619	1173	4641	0.000	0.019	0.02	Malignant neoplasm of prostate
C3	I350	2664	4641	0.001	0.043	0.03	Aortic (valve) stenosis
C3	M171	2940	4641	0.002	0.047	0.03	Other primary gonarthrosis
C3	J441	1678	4641	0.001	0.027	0.03	Chronic obstructive pulmonary disease with acute exacerbation, unspecified
C3	N409	3319	4641	0.002	0.053	0.04	Hyperplasia of prostate
C3	I509	6160	4641	0.004	0.098	0.04	Heart failure, unspecified
C3	M059	682	4641	0.000	0.011	0.04	Seropositive rheumatoid arthritis, unspecified
C4	N180	454	4401	0.002	0.007	0.26	Chronic kidney disease
C4	J350	315	4401	0.001	0.005	0.28	Chronic tonsillitis
C4	E105	416	4401	0.002	0.007	0.28	Type 1 diabetes mellitus: With peripheral circulatory complications
C4	E103	576	4401	0.003	0.009	0.30	Type 1 diabetes mellitus: With ophthalmic complications
C4	K801	411	4401	0.002	0.006	0.36	Calculus of gallbladder with other cholecystitis
C4	E107	583	4401	0.003	0.009	0.38	Type 1 diabetes mellitus: With multiple complications
C4	M235	269	4401	0.002	0.004	0.38	Chronic instability of knee

Clst	ICD-10	# code	# clst	Obs	Exp	O/E-ratio	description
C4	G439	463	4401	0.003	0.007	0.41	Migraine, unspecified
C4	H360	1467	4401	0.010	0.023	0.43	Diabetic retinopathy
C4	J039	402	4401	0.003	0.006	0.44	Acute tonsillitis, unspecified
C5	D303	600	4290	0.003	0.009	0.27	Benign neoplasm: Bladder
C5	I495	482	4290	0.002	0.008	0.31	Sick sinus syndrome
C5	K402	280	4290	0.001	0.004	0.32	Bilateral inguinal hernia, without obstruction or gangrene
C5	I491	273	4290	0.001	0.004	0.33	Atrial premature depolarization
C5	K409	3787	4290	0.021	0.059	0.35	Unilateral or unspecified inguinal hernia, without obstruction or gangrene
C5	M235	269	4290	0.002	0.004	0.39	Chronic instability of knee
C5	M539	274	4290	0.002	0.004	0.44	Dorsopathy, unspecified
C5	I479	614	4290	0.004	0.009	0.44	Paroxysmal tachycardia, unspecified
C5	C509	1103	4290	0.008	0.017	0.45	Malignant neoplasm: Breast, unspecified
C5	N508	300	4290	0.002	0.005	0.45	Other specified disorders of male genital organs
C6	N180	454	3589	0.001	0.007	0.12	Chronic kidney disease
C6	C679	292	3589	0.001	0.005	0.12	Malignant neoplasm: Bladder, unspecified
C6	N199	711	3589	0.001	0.011	0.12	Unspecified kidney failure
C6	I489	7075	3589	0.015	0.110	0.14	Atrial fibrillation and atrial flutter, unspecified
C6	J441	1678	3589	0.004	0.026	0.14	Chronic obstructive pulmonary disease with acute exacerbation, unspecified
C6	J180	257	3589	0.001	0.004	0.14	Bronchopneumonia, unspecified
C6	C509	1103	3589	0.003	0.017	0.15	Malignant neoplasm: Breast, unspecified
C6	J960	837	3589	0.002	0.013	0.15	Acute respiratory failure
C6	J440	743	3589	0.002	0.012	0.17	Chronic obstructive pulmonary disease with acute lower respiratory infection
C6	K409	3787	3589	0.010	0.059	0.17	Unilateral or unspecified inguinal hernia, without obstruction or gangrene
C7	E107	583	3309	0.000	0.009	0.03	Type 1 diabetes mellitus: With multiple complications
C7	E103	576	3309	0.000	0.009	0.03	Type 1 diabetes mellitus: With ophthalmic complications
C7	N180	454	3309	0.000	0.007	0.04	Chronic kidney disease
C7	E104	434	3309	0.000	0.007	0.04	Type 1 diabetes mellitus: With neurological complications
C7	I709	433	3309	0.000	0.007	0.04	Generalized and unspecified atherosclerosis
C7	E105	416	3309	0.000	0.007	0.05	Type 1 diabetes mellitus: With peripheral circulatory complications
C7	E102	377	3309	0.000	0.006	0.05	Type 1 diabetes mellitus: With renal complications
C7	N189	1094	3309	0.001	0.017	0.05	Chronic kidney disease, unspecified
C7	J180	257	3309	0.000	0.004	0.07	Bronchopneumonia, unspecified
C7	E115	551	3309	0.001	0.009	0.11	Type 2 diabetes mellitus: With peripheral circulatory complications
C8	E103	576	2802	0.000	0.009	0.04	Type 1 diabetes mellitus: With ophthalmic complications
C8	E108	1073	2802	0.001	0.017	0.09	Type 1 diabetes mellitus: With unspecified complications
C8	E104	434	2802	0.001	0.007	0.11	Type 1 diabetes mellitus: With neurological complications
C8	E105	416	2802	0.001	0.006	0.11	Type 1 diabetes mellitus: With peripheral circulatory complications

Clst	ICD-10	# code	# clst	Obs	Exp	O/E-ratio	description
C8	E115	551	2802	0.001	0.009	0.13	Type 2 diabetes mellitus: With peripheral circulatory complications
C8	M512	366	2802	0.001	0.006	0.13	Other specified intervertebral disc displacement
C8	E162	510	2802	0.001	0.008	0.14	Hypoglycaemia, unspecified
C8	E109	2680	2802	0.006	0.041	0.15	Type 1 diabetes mellitus: Without complications
C8	N180	454	2802	0.001	0.007	0.15	Chronic kidney disease
C8	E107	583	2802	0.001	0.009	0.16	Type 1 diabetes mellitus: With multiple complications
C9	M539	274	2581	0.000	0.004	0.09	Dorsopathy, unspecified
C9	D179	257	2581	0.000	0.004	0.10	Benign lipomatous neoplasm, unspecified
C9	D251	256	2581	0.000	0.004	0.10	Intramural leiomyoma of uterus
C9	D303	600	2581	0.001	0.009	0.13	Benign neoplasm: Bladder
C9	R072	367	2581	0.001	0.006	0.14	Precordial pain
C9	M512	366	2581	0.001	0.006	0.14	Other specified intervertebral disc displacement
C9	G439	463	2581	0.001	0.007	0.16	Migraine, unspecified
C9	M653	738	2581	0.002	0.011	0.17	Trigger finger
C9	M511	3357	2581	0.009	0.052	0.17	Lumbar and other intervertebral disc disorders with radiculopathy
C9	M431	583	2581	0.002	0.009	0.17	Spondylolisthesis
C10	A630	320	2562	0.000	0.005	0.08	Anogenital (venereal) warts
C10	K298	256	2562	0.000	0.004	0.10	Duodenitis
C10	I999	383	2562	0.001	0.006	0.13	Other and unspecified disorders of circulatory system
C10	F172	349	2562	0.001	0.005	0.14	Mental and behavioural disorders due to use of tobacco: Dependence syndrome
C10	M771	492	2562	0.001	0.008	0.16	Lateral epicondylitis
C10	N479	462	2562	0.001	0.007	0.16	Redundant prepuce, phimosis and paraphimosis
C10	N180	454	2562	0.001	0.007	0.17	Chronic kidney disease
C10	I309	297	2562	0.001	0.005	0.17	Acute pericarditis, unspecified
C10	I830	293	2562	0.001	0.005	0.17	Varicose veins of lower extremities with ulcer
C10	I802	280	2562	0.001	0.004	0.18	Phlebitis and thrombophlebitis of other deep vessels of lower extremities
C11	C509	1103	2292	0.000	0.017	0.03	Malignant neoplasm: Breast, unspecified
C11	H360	1467	2292	0.001	0.023	0.04	Diabetic retinopathy
C11	E113	720	2292	0.000	0.011	0.04	Type 2 diabetes mellitus: With ophthalmic complications
C11	E107	583	2292	0.000	0.009	0.05	Type 1 diabetes mellitus: With multiple complications
C11	E103	576	2292	0.000	0.009	0.05	Type 1 diabetes mellitus: With ophthalmic complications
C11	E104	434	2292	0.000	0.007	0.06	Type 1 diabetes mellitus: With neurological complications
C11	E105	416	2292	0.000	0.006	0.07	Type 1 diabetes mellitus: With peripheral circulatory complications
C11	E112	761	2292	0.001	0.012	0.07	Type 2 diabetes mellitus: With renal complications
C11	D249	758	2292	0.001	0.012	0.07	Benign neoplasm of breast
C11	E108	1073	2292	0.001	0.017	0.08	Type 1 diabetes mellitus: With unspecified complications
C12	C509	1103	2213	0.000	0.017	0.03	Malignant neoplasm: Breast, unspecified
C12	N921	875	2213	0.000	0.013	0.03	Excessive and frequent menstruation with irregular cycle
C12	E103	576	2213	0.000	0.009	0.05	Type 1 diabetes mellitus: With ophthalmic complications
C12	N924	561	2213	0.000	0.009	0.05	Excessive bleeding in the premenopausal period

Clst	ICD-10	# code	# clst	Obs	Exp	O/E-ratio	description
C12	E148	483	2213	0.000	0.007	0.06	Unspecified diabetes mellitus: With unspecified complications
C12	N920	948	2213	0.001	0.015	0.06	Excessive and frequent menstruation with regular cycle
C12	F419	335	2213	0.000	0.005	0.09	Anxiety disorder, unspecified
C12	R064	542	2213	0.001	0.008	0.11	Hyperventilation
C12	D251	256	2213	0.000	0.004	0.12	Intramural leiomyoma of uterus
C12	N832	489	2213	0.001	0.008	0.12	Other and unspecified ovarian cysts
C13	E112	761	2070	0.000	0.012	0.04	Type 2 diabetes mellitus: With renal complications
C13	E102	377	2070	0.000	0.006	0.08	Type 1 diabetes mellitus: With renal complications
C13	N179	377	2070	0.000	0.006	0.08	Acute renal failure, unspecified
C13	D509	459	2070	0.001	0.007	0.14	Iron deficiency anaemia, unspecified
C13	H330	418	2070	0.001	0.006	0.15	Retinal detachment with retinal break
C13	J819	545	2070	0.001	0.008	0.17	NA
C13	E162	510	2070	0.001	0.008	0.19	Hypoglycaemia, unspecified
C13	E117	980	2070	0.003	0.015	0.19	Type 2 diabetes mellitus: With multiple complications
C13	I501	1502	2070	0.005	0.023	0.21	Left ventricular failure
C13	C679	292	2070	0.001	0.004	0.22	Malignant neoplasm: Bladder, unspecified
C14	H360	1467	2040	0.002	0.022	0.11	Diabetic retinopathy
C14	E103	576	2040	0.001	0.009	0.17	Type 1 diabetes mellitus: With ophthalmic complications
C14	E105	416	2040	0.001	0.006	0.23	Type 1 diabetes mellitus: With peripheral circulatory complications
C14	G510	365	2040	0.001	0.006	0.26	Bell's palsy
C14	H431	334	2040	0.001	0.005	0.29	Vitreous haemorrhage
C14	E113	720	2040	0.003	0.011	0.31	Type 2 diabetes mellitus: With ophthalmic complications
C14	M519	604	2040	0.003	0.009	0.32	Intervertebral disc disorder, unspecified
C14	L905	295	2040	0.001	0.004	0.33	Scar conditions and fibrosis of skin
C14	E107	583	2040	0.003	0.009	0.33	Type 1 diabetes mellitus: With multiple complications
C14	H438	291	2040	0.001	0.004	0.33	Other disorders of vitreous body
C15	J439	317	2013	0.000	0.005	0.10	Emphysema, unspecified
C15	C679	292	2013	0.000	0.004	0.11	Malignant neoplasm: Bladder, unspecified
C15	K510	283	2013	0.000	0.004	0.12	Ulcerative (chronic) pancolitis
C15	I495	482	2013	0.001	0.007	0.14	Sick sinus syndrome
C15	K045	333	2013	0.001	0.005	0.20	Chronic apical periodontitis
C15	R570	320	2013	0.001	0.005	0.20	Cardiogenic shock
C15	R590	305	2013	0.001	0.005	0.21	Localized enlarged lymph nodes
C15	N180	454	2013	0.001	0.007	0.22	Chronic kidney disease
C15	D303	600	2013	0.002	0.009	0.22	Benign neoplasm: Bladder
C15	N133	267	2013	0.001	0.004	0.24	Other and unspecified hydronephrosis
C16	E149	958	1654	0.001	0.015	0.04	Unspecified diabetes mellitus: Without complications
C16	E114	881	1654	0.001	0.013	0.04	Type 2 diabetes mellitus: With neurological complications
C16	E103	576	1654	0.001	0.009	0.07	Type 1 diabetes mellitus: With ophthalmic complications
C16	H360	1467	1654	0.002	0.022	0.08	Diabetic retinopathy
C16	E105	416	1654	0.001	0.006	0.10	Type 1 diabetes mellitus: With peripheral circulatory complications
C16	E102	377	1654	0.001	0.006	0.10	Type 1 diabetes mellitus: With renal complications

Clst	ICD-10	# code	# clst	Obs	Exp	O/E-ratio	description
C16	E108	1073	1654	0.002	0.016	0.11	Type 1 diabetes mellitus: With unspecified complications
C16	H431	334	1654	0.001	0.005	0.12	Vitreous haemorrhage
C16	E117	980	1654	0.002	0.015	0.12	Type 2 diabetes mellitus: With multiple complications
C16	E107	583	1654	0.001	0.009	0.14	Type 1 diabetes mellitus: With multiple complications
C17	N199	711	1281	0.001	0.011	0.07	Unspecified kidney failure
C17	E162	510	1281	0.001	0.008	0.10	Hypoglycaemia, unspecified
C17	E105	416	1281	0.001	0.006	0.12	Type 1 diabetes mellitus: With peripheral circulatory complications
C17	I469	372	1281	0.001	0.006	0.14	Cardiac arrest, unspecified
C17	D303	600	1281	0.002	0.009	0.17	Benign neoplasm: Bladder
C17	E107	583	1281	0.002	0.009	0.18	Type 1 diabetes mellitus: With multiple complications
C17	E103	576	1281	0.002	0.009	0.18	Type 1 diabetes mellitus: With ophthalmic complications
C17	K580	576	1281	0.002	0.009	0.18	Irritable bowel syndrome with diarrhoea
C17	K650	264	1281	0.001	0.004	0.20	Acute peritonitis
C17	D179	257	1281	0.001	0.004	0.20	Benign lipomatous neoplasm, unspecified
C18	J450	401	1251	0.001	0.006	0.13	Predominantly allergic asthma
C18	H521	366	1251	0.001	0.006	0.14	Myopia
C18	J350	315	1251	0.001	0.005	0.17	Chronic tonsillitis
C18	N924	561	1251	0.002	0.008	0.19	Excessive bleeding in the premenopausal period
C18	K402	280	1251	0.001	0.004	0.19	Bilateral inguinal hernia, without obstruction or gangrene
C18	N840	546	1251	0.002	0.008	0.19	Polyp of corpus uteri
C18	M235	269	1251	0.001	0.004	0.20	Chronic instability of knee
C18	M234	258	1251	0.001	0.004	0.20	Loose body in knee
C18	N832	489	1251	0.002	0.007	0.22	Other and unspecified ovarian cysts
C18	K800	468	1251	0.002	0.007	0.23	Calculus of gallbladder with acute cholecystitis
C19	C509	1103	1168	0.002	0.017	0.10	Malignant neoplasm: Breast, unspecified
C19	D649	1637	1168	0.003	0.025	0.10	Anaemia, unspecified
C19	F103	518	1168	0.001	0.008	0.11	Mental and behavioural disorders due to use of alcohol: Withdrawal state
C19	K810	505	1168	0.001	0.008	0.11	Acute cholecystitis
C19	N811	974	1168	0.002	0.015	0.12	Cystocele
C19	I429	479	1168	0.001	0.007	0.12	Cardiomyopathy, unspecified
C19	I709	433	1168	0.001	0.007	0.13	Generalized and unspecified atherosclerosis
C19	N390	1219	1168	0.003	0.018	0.14	Urinary tract infection, site not specified
C19	R001	394	1168	0.001	0.006	0.14	Bradycardia, unspecified
C19	E871	393	1168	0.001	0.006	0.14	Hypo-osmolality and hyponatraemia
C20	N920	948	1119	0.001	0.014	0.06	Excessive and frequent menstruation with regular cycle
C20	N921	875	1119	0.001	0.013	0.07	Excessive and frequent menstruation with irregular cycle
C20	G442	667	1119	0.001	0.010	0.09	Tension-type headache
C20	N924	561	1119	0.001	0.008	0.10	Excessive bleeding in the premenopausal period
C20	D279	435	1119	0.001	0.007	0.14	NA
C20	E104	434	1119	0.001	0.007	0.14	Type 1 diabetes mellitus: With neurological complications
C20	M224	423	1119	0.001	0.006	0.14	Chondromalacia patellae
C20	R104	391	1119	0.001	0.006	0.15	Other and unspecified abdominal pain
C20	E041	369	1119	0.001	0.006	0.16	Nontoxic single thyroid nodule

Clst	ICD-10	# code	# clst	Obs	Exp	O/E-ratio	description
C20	M512	366	1119	0.001	0.006	0.16	Other specified intervertebral disc displacement
C21	J960	837	1000	0.001	0.013	0.08	Acute respiratory failure
C21	E107	583	1000	0.001	0.009	0.11	Type 1 diabetes mellitus: With multiple complications
C21	E103	576	1000	0.001	0.009	0.12	Type 1 diabetes mellitus: With ophthalmic complications
C21	I493	575	1000	0.001	0.009	0.12	Ventricular premature depolarization
C21	I495	482	1000	0.001	0.007	0.14	Sick sinus syndrome
C21	N180	454	1000	0.001	0.007	0.15	Chronic kidney disease
C21	R091	418	1000	0.001	0.006	0.16	Pleurisy
C21	E789	417	1000	0.001	0.006	0.16	Disorder of lipoprotein metabolism, unspecified
C21	E871	393	1000	0.001	0.006	0.17	Hypo-osmolality and hyponatraemia
C21	E102	377	1000	0.001	0.006	0.18	Type 1 diabetes mellitus: With renal complications
C22	F100	1212	988	0.001	0.018	0.06	Mental and behavioural disorders due to use of alcohol: Acute intoxication
C22	L979	605	988	0.001	0.009	0.11	Ulcer of lower limb, not elsewhere classified
C22	I429	479	988	0.001	0.007	0.14	Cardiomyopathy, unspecified
C22	N180	454	988	0.001	0.007	0.15	Chronic kidney disease
C22	M224	423	988	0.001	0.006	0.16	Chondromalacia patellae
C22	J448	396	988	0.001	0.006	0.17	Other specified chronic obstructive pulmonary disease
C22	E102	377	988	0.001	0.006	0.18	Type 1 diabetes mellitus: With renal complications
C22	I469	372	988	0.001	0.006	0.18	Cardiac arrest, unspecified
C22	R570	320	988	0.001	0.005	0.21	Cardiogenic shock
C22	L400	319	988	0.001	0.005	0.21	Psoriasis vulgaris
C23	F103	518	935	0.001	0.008	0.14	Mental and behavioural disorders due to use of alcohol: Withdrawal state
C23	M191	474	935	0.001	0.007	0.15	Post-traumatic arthrosis of other joints
C23	R490	415	935	0.001	0.006	0.17	Dysphonia
C23	C619	1173	935	0.003	0.018	0.18	Malignant neoplasm of prostate
C23	I999	383	935	0.001	0.006	0.18	Other and unspecified disorders of circulatory system
C23	E041	369	935	0.001	0.006	0.19	Nontoxic single thyroid nodule
C23	H659	368	935	0.001	0.006	0.19	Nonsuppurative otitis media, unspecified
C23	I480	364	935	0.001	0.005	0.20	Paroxysmal atrial fibrillation
C23	H908	631	935	0.002	0.010	0.22	Mixed conductive and sensorineural hearing loss, unspecified
C23	I839	1522	935	0.005	0.023	0.23	Varicose veins of lower extremities without ulcer or inflammation
C24	N409	3319	932	0.002	0.050	0.04	Hyperplasia of prostate
C24	C619	1173	932	0.001	0.018	0.06	Malignant neoplasm of prostate
C24	E108	1073	932	0.001	0.016	0.07	Type 1 diabetes mellitus: With unspecified complications
C24	H833	1412	932	0.002	0.021	0.10	Noise effects on inner ear
C24	E107	583	932	0.001	0.009	0.12	Type 1 diabetes mellitus: With multiple complications
C24	M109	547	932	0.001	0.008	0.13	Gout, unspecified
C24	E116	477	932	0.001	0.007	0.15	Type 2 diabetes mellitus: With other specified complications
C24	M191	474	932	0.001	0.007	0.15	Post-traumatic arthrosis of other joints
C24	H109	465	932	0.001	0.007	0.15	Conjunctivitis, unspecified
C24	N484	456	932	0.001	0.007	0.16	Impotence of organic origin
C25	H938	1116	860	0.001	0.017	0.07	Other specified disorders of ear
C25	H360	1467	860	0.002	0.022	0.10	Diabetic retinopathy
C25	N200	1391	860	0.002	0.021	0.11	Calculus of kidney

Clst	ICD-10	# code	# clst	Obs	Exp	O/E-ratio	description
C25	H911	3527	860	0.007	0.053	0.13	Presbycusis
C25	C619	1173	860	0.002	0.018	0.13	Malignant neoplasm of prostate
C25	E107	583	860	0.001	0.009	0.13	Type 1 diabetes mellitus: With multiple complications
C25	E103	576	860	0.001	0.009	0.13	Type 1 diabetes mellitus: With ophthalmic complications
C25	I493	575	860	0.001	0.009	0.13	Ventricular premature depolarization
C25	H919	4610	860	0.010	0.069	0.15	Hearing loss, unspecified
C25	H833	1412	860	0.003	0.021	0.16	Noise effects on inner ear
C26	E112	761	852	0.001	0.011	0.10	Type 2 diabetes mellitus: With renal complications
C26	N409	3319	852	0.006	0.050	0.12	Hyperplasia of prostate
C26	C619	1173	852	0.002	0.018	0.13	Malignant neoplasm of prostate
C26	I350	2664	852	0.006	0.040	0.15	Aortic (valve) stenosis
C26	I714	517	852	0.001	0.008	0.15	Abdominal aortic aneurysm, without mention of rupture
C26	E162	510	852	0.001	0.008	0.15	Hypoglycaemia, unspecified
C26	H360	1467	852	0.004	0.022	0.16	Diabetic retinopathy
C26	I495	482	852	0.001	0.007	0.16	Sick sinus syndrome
C26	K800	468	852	0.001	0.007	0.17	Calculus of gallbladder with acute cholecystitis
C26	N180	454	852	0.001	0.007	0.17	Chronic kidney disease
C27	H360	1467	823	0.001	0.022	0.06	Diabetic retinopathy
C27	M170	2145	823	0.004	0.032	0.11	Primary gonarthrosis, bilateral
C27	D303	600	823	0.001	0.009	0.14	Benign neoplasm: Bladder
C27	C509	1103	823	0.002	0.017	0.15	Malignant neoplasm: Breast, unspecified
C27	J209	527	823	0.001	0.008	0.15	Acute bronchitis, unspecified
C27	J320	513	823	0.001	0.008	0.16	Chronic maxillary sinusitis
C27	E162	510	823	0.001	0.008	0.16	Hypoglycaemia, unspecified
C27	L022	506	823	0.001	0.008	0.16	Cutaneous abscess, furuncle and carbuncle of trunk
C27	K810	505	823	0.001	0.008	0.16	Acute cholecystitis
C27	M179	2242	823	0.006	0.034	0.18	Gonarthrosis, unspecified
C28	M546	977	686	0.001	0.015	0.10	Pain in thoracic spine
C28	D249	758	686	0.001	0.011	0.13	Benign neoplasm of breast
C28	E113	720	686	0.001	0.011	0.14	Type 2 diabetes mellitus: With ophthalmic complications
C28	E669	2114	686	0.004	0.032	0.14	Obesity, unspecified
C28	R002	691	686	0.001	0.010	0.14	Palpitations
C28	K359	678	686	0.001	0.010	0.14	NA
C28	H350	548	686	0.001	0.008	0.18	Background retinopathy and retinal vascular changes
C28	I119	534	686	0.001	0.008	0.18	Hypertensive heart disease without (congestive) heart failure
C28	J209	527	686	0.001	0.008	0.18	Acute bronchitis, unspecified
C28	E162	510	686	0.001	0.008	0.19	Hypoglycaemia, unspecified
C29	M171	2940	550	0.002	0.044	0.04	Other primary gonarthrosis
C29	I489	7075	550	0.005	0.106	0.05	Atrial fibrillation and atrial flutter, unspecified
C29	H919	4610	550	0.004	0.069	0.05	Hearing loss, unspecified
C29	I702	2251	550	0.002	0.034	0.05	Atherosclerosis of arteries of extremities
C29	M179	2242	550	0.002	0.034	0.05	Gonarthrosis, unspecified
C29	M170	2145	550	0.002	0.032	0.06	Primary gonarthrosis, bilateral
C29	G459	2066	550	0.002	0.031	0.06	Transient cerebral ischaemic attack, unspecified
C29	I639	1989	550	0.002	0.030	0.06	Cerebral infarction, unspecified
C29	G473	1897	550	0.002	0.028	0.06	Sleep apnoea
C29	N409	3319	550	0.004	0.050	0.07	Hyperplasia of prostate

Clst	ICD-10	# code	# clst	Obs	Exp	O/E-ratio	description
C30	F100	1212	533	0.002	0.018	0.10	Mental and behavioural disorders due to use of alcohol: Acute intoxication
C30	N393	827	533	0.002	0.012	0.15	Stress incontinence
C30	M503	520	533	0.002	0.008	0.24	Other cervical disc degeneration
C30	R065	983	533	0.004	0.015	0.26	Mouth breathing
C30	M546	977	533	0.004	0.015	0.26	Pain in thoracic spine
C30	R490	415	533	0.002	0.006	0.30	Dysphonia
C30	M060	398	533	0.002	0.006	0.32	Seronegative rheumatoid arthritis
C30	F102	1189	533	0.006	0.018	0.32	Mental and behavioural disorders due to use of alcohol: Dependence syndrome
C30	J448	396	533	0.002	0.006	0.32	Other specified chronic obstructive pulmonary disease
C30	C509	1103	533	0.006	0.017	0.34	Malignant neoplasm: Breast, unspecified
C31	F102	1189	520	0.002	0.018	0.11	Mental and behavioural disorders due to use of alcohol: Dependence syndrome
C31	C619	1173	520	0.002	0.018	0.11	Malignant neoplasm of prostate
C31	K439	981	520	0.002	0.015	0.13	Other and unspecified ventral hernia without obstruction or gangrene
C31	E117	980	520	0.002	0.015	0.13	Type 2 diabetes mellitus: With multiple complications
C31	E112	761	520	0.002	0.011	0.17	Type 2 diabetes mellitus: With renal complications
C31	I351	696	520	0.002	0.010	0.18	Aortic (valve) insufficiency
C31	N200	1391	520	0.004	0.021	0.18	Calculus of kidney
C31	E118	2669	520	0.008	0.040	0.19	Type 2 diabetes mellitus: With unspecified complications
C31	M519	604	520	0.002	0.009	0.21	Intervertebral disc disorder, unspecified
C31	E107	583	520	0.002	0.009	0.22	Type 1 diabetes mellitus: With multiple complications

S6 Table: Chi-squared test for distribution laboratory values in clusters

Component	P-val.	Adj. P-val.
Alanine transaminase (ALAT)	4.78 e-22	1.15e-20
Albumin	4.81e-22	1.15e-20
Alkaline phosphatase	2.01e-22	4.82e-21
Bilirubin	1.09e-13	2.60e-12
C-reactive protein (CRP)	1.65e-96	3.95e-95
Carbamide	5.49-e200	1.32e-198
Cholesterol HDL	1.99e-66	4.77e-65
Cholesterol LDL	4.86e-53	1.17e-51
Cholesterol total	2.64e-58	6.34 e-57
Coagulation factor II + VII + X	7.96e-280	1.91e-278
Creatinine	9.28e-302	2.23e-300
Eosinophils	4.43e-6	1.06e-4
Estimated glomerular filtration rate (eGFR)	0	0
Glucose	0	0
Hemoglobin	2.77e-218	6.65e-217
Leukocytes	1.42e-39	3.41e-38
Lymphocytes	1.54e-17	3.69e-16
Monocytes	1.06e-11	2.55e-10
Neutrophils	5.69e-20	1.36e-18
Platelets	2.39e-23	5.73e-22
Potassium	9.03e-32	2.17e-30
Sodium	2.24e-74	5.38e-74
Triglyceride	2.10e-60	5.04 e-59
Troponin	7.10e-73	1.70e-71

S7 Table: Traits with significantly different PGS distributions in clusters

Cluster	n	trait	effect	effect size	FDR
C1	2,025	Systolic Blood Pressure	+	0.20	<0.0005
		Diastolic Blood Pressure	+	0.16	<0.0005
		Total Cholesterol	-	-0.08	0.026
C4	1,532	Atrial Fibrillation	+	0.57	<0.0005
		Heart Failure	+	0.08	0.031
		Coronary Artery Disease	-	-0.12	0.001
		T2D (BMI-adj.)	-	-0.11	0.001
		Acute Myocardial Infarction	-	-0.08	0.031
		Triglyceride	-	-0.08	0.044
		Total Cholesterol	-	-0.08	0.046
C5	1,136	T2D (BMI-adj.)	+	0.55	<0.0005
		NAFLD	+	0.11	0.021
C6	860	Total Cholesterol	+	0.21	<0.0005
		Triglyceride	+	0.20	<0.0005
		LDL Cholesterol	+	0.15	0.001
		Coronary Artery Disease	+	0.15	0.001
		Diastolic Blood Pressure	-	-0.13	0.015
		Systolic Blood Pressure	-	-0.11	0.040
C8	817	Systolic Blood Pressure	-	-0.16	0.001
		Stroke	-	-0.12	0.023
		Coronary Artery Disease	-	-0.12	0.028
		Diastolic Blood Pressure	-	-0.11	0.031
		LDL Cholesterol	-	-0.10	0.047
C10	744	Coronary Artery Disease	-	-0.13	0.017
		Acute Myocardial Infarction	-	-0.13	0.021
C11	718	Stroke	-	-0.11	0.040
		Heart Failure	-	-0.11	0.040
C12	649	Coronary Artery Disease	-	-0.14	0.013
		Acute Myocardial Infarction	-	-0.12	0.033
C13	606	Diastolic Blood Pressure	-	-0.12	0.040
C15	588	LDL Cholesterol	+	0.14	0.020
		Total Cholesterol	+	0.14	0.026
C17	348	Systolic Blood Pressure	+	0.16	0.040
C18	481	T2D (BMI-adj.)	+	0.15	0.023
		Acute Myocardial Infarction	+	0.15	0.028
		Atrial Fibrillation	-	-0.13	0.049
C23	290	T2D (BMI-adj.)	+	0.27	0.001
C25	297	Coronary Artery Disease	+	0.24	0.002
		Acute Myocardial Infarction	+	0.18	0.032
		Heart Failure	+	0.18	0.031
C27	231	Stroke	+	0.24	0.015

Supplementary Material

Manuscript title: Subgrouping multimorbid patients with ischemic heart disease by means of unsupervised clustering: A cohort study of 72,249 patients defined comprehensively by diagnoses prior to presentation

Authors: Haue AD, Holm PC, et al.

S1 Appendix: Construction of patient similarity network, MCL algorithm settings and assessment of cluster robustness

To define the patient similarity network, a lower rank approximation of the *n* x *m* matrix was created using the "truncatedSVD" implecpmentation of SVD from the python package scikit-learn with 41 components, 10 iterations, and fixed random seed of 42 to ensure reproducible results. Thus, the 3,046 diagnoses were represented in 41 components based on this lower rank approximation. By selecting 41 components, the accumulated explained variance ratio was 0.50 (S2 Fig). To reduce the density of the patient similarity network, while still retaining an informative topology, all edges with an edge weight less than 0.3 were removed and the number of edges connected to each node were limited using the "#ceilnb" transformation from "mcl-edge"; a maximum of 8000 was used (S3 Fig). The weights of the remaining edges in the network were shifted such that the lowest weight was 0.0 as recommended in the MCL manual. The final pre-processed networks were then used as input for the "mcl" implementation of the MCL algorithm [1]. We used a pruning scheme of -P 7000 -S 800 -R 900 -pct 90, and a pre-inflation factor of 0.5 to make edge-weights more homogenous. For the MCL clustering, we selected a pre-inflation parameter of 2.0 corresponding to the default in the MCL manual [2].

For cluster robustness assessment, diluted versions of the reference clustering were generated by deleting edges with a probability of α, where α would range between 0 and 50% [3]. An α of 0 would leave the network unchanged. In contrast, the shuffled versions of the network had the same number of nodes and edges as the reference clustering. The shuffled networks were generated as described by Karrer et al. [4]. Finally, the generated clusters were compared to the reference clustering and the variances were quantified with reference to the so-called variation of information measure (VI) [5].

S2 Appendix: Preprocessing of laboratory data

The laboratory test results from the EHR data were originally archived in the administrative biochemical databases Labka and BCC [6]. In relation to the EHR data used in this study, Labka covers the hospitals with SHAK codes 1301, 1309, 1330, 1351, 1401, 1501, 1516 and 4001 in the Capital Region of Denmark and BCC covers the hospitals with SHAK codes 2000 and 2501 in Region Zealand for the periods 2009-2016 and 2012-2016, respectively (S2 Table). Biochemical laboratory tests were either classified in accordance with the Nomenclature, Properties and Units (NPU) or local systems [7]. Reference intervals were provided from the laboratories that analyzed the blood tests. Biochemical data was expected to be available for the patients where the index procedure was performed at a hospital located in either the Capital Region or Region Zealand at a time that was covered by the two databases.

A total of 48,957 patients (30,736 males and 18,221 females) were included from a hospital where biochemical data was available (67.8% of the entire cohort). As an indicator for data completeness and quality, the number of patients where available biochemical data at time of index agreed with the clinical standard of care was assessed. This implied that patients had sodium, potassium, hemoglobin, and creatinine (or estimated glomerular filtration rate) measured maximum 90 days before or at the day of index. The 31,224 patients who fulfilled this requirement were included in the biochemical analysis. Laboratory measurements available for at least 50% of these patients were included in the analysis. In cases where patients had more than one test available in the period up from 90 days before to index, the test closest to index was used. As listed in the main text included samples were plasma levels of potassium, sodium, hemoglobin, estimated glomerular filtration (eGFR), creatinine, carbamide, glucose, troponin (I/T), HDL cholesterol, LDL cholesterol, total cholesterol, leukocytes, C-reactive protein, lymphocytes, monocytes, neutrophils, basophils,

platelets, INR, alanine transaminase, albumin, alkaline phosphatase, bilirubin, and triglyceride. All analyses of biochemical data were performed in R 3.6.2 using the “ComplexHeatmap” and “circlize” packages .

S3 Appendix: Calculation of polygenetic risk scores for 14 traits

Polygenic risk scores were calculated using the LDpred2 framework, implemented in the R package bigsnpr (v1.11.6) with R version 4.0.0 and the workflow management system Snakemake [11–13]. In preparation for PGS calculations, autosomal genotype data from 242,644 individuals in the Copenhagen Hospital Biobank – Cardiovascular Disease Cohort (CHB-CVDC) [14] was filtered to only include variants present in LDpred2's recommended set of 1,054,330 reference variants. This recommended set is based on the reference set HapMap3 from the International HapMap project, which was established by genotyping 1.6 million single nucleotide polymorphisms (SNPs) in 1,184 individuals from 11 global populations [15]. Any missing genotype information was assumed to be the affected locus' reference allele.

We matched the remaining set of 994,643 genotyped variants with variants found in summary statistics data corresponding to 14 traits, obtained from nine GWAS meta-analyses (atrial fibrillation [16], BMI-adjusted type 2 diabetes [17], chronic kidney disease [18], HDL cholesterol levels [19], heart failure [20], LDL cholesterol levels [19], stroke [21], total cholesterol levels [19], triglyceride levels [19]) and five GWAS (acute myocardial infarction [22], coronary artery disease [23], diastolic blood pressure [24], non-alcoholic fatty liver disease [25], systolic blood pressure [24]). Variants present in both genotype and summary statistics data were then subject to LDpred2's recommended standard deviation quality control. After variant matching and quality control, a mean of 963,354 (S.D. 87,774) variants remained for subsequent per-chromosome risk score calculation for each of the 14 traits. We used the LDpred2-auto algorithm with 30 Gibbs sampling chains, 1,000 burn-in iterations and 500 iterations after burn-in. The initial values for the 30 sampling chains were a) the LDSC regression estimate for heritability h^2 (same for all chains); b) one of 30 initial values for the proportion of causal variants p, evenly spaced on a logarithmic scale from 10^{-4} to 0.5.

Variant effect sizes were calculated from each set of 30 sampling chains (per trait and chromosome) through a three-step process, which serves to ensure that the model (spanning 30 chains) successfully converged: 1) computing the standard deviations of each chains' predicted scores, 2) keeping only the chains within three median absolute deviations from the median standard deviation, 3) averaging the effect sizes of the remaining chains. Across the 308 per-chromosome models (14 traits times 22 chromosomes), 28.9 chains were included in the final score on average. For each individual, we calculated per-chromosome risk scores by multiplying the average variant effect sizes with the individual's corresponding genotype, and then added the per-chromosome risk scores up into one genome-wide PGS. To ease comparisons across traits, each trait's PGS distribution was scaled to a mean of zero and a standard deviation of one.

References, S1-3 Appendices

1. Van Dongen, S. M. Graph clustering by flow simulation. (2000).
2. MCL - a cluster algorithm for graphs. http://micans.org/mcl/.
3. Kirk, I. K. *et al.* Linking glycemic dysregulation in diabetes to symptoms, comorbidities, and genetics through EHR data mining. *eLife* **8**, e44941 (2019).
4. Karrer, B., Levina, E. & Newman, M. E. J. Robustness of community structure in networks. *Phys. Rev. E* **77**, 046119 (2008).
5. Meilă, M. Comparing clusterings—an information based distance. *J. Multivar. Anal.* **98**, 873–895 (2007).
6. Grann, A. F., Erichsen, R., Nielsen, A. G., Frøslev, T. & Thomsen, R. W. Existing data sources for clinical epidemiology: The clinical laboratory information system (LABKA) research database at Aarhus University, Denmark. *Clin. Epidemiol.* **3**, 133–138 (2011).
7. Petersen, U. M., Dybkaer, R. & Olesen, H. Properties and units in the clinical laboratory sciences. Part XXIII. The NPU terminology, principles, and implementation: A user's guide (IUPAC Technical Report)*. *Pure Appl Chem* **84**, 137–165 (2012).
8. R Core Team. *R: A Language and Environment for Statistical Computing*. (R Foundation for Statistical Computing, 2019).
9. Gu, Z., Gu, L., Eils, R., Schlesner, M. & Brors, B. circlize implements and enhances circular visualization in R. *Bioinformatics* **30**, 2811–2812 (2014).
10. Gu, Z., Eils, R. & Schlesner, M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* **32**, 2847–2849 (2016).
11. Mölder, F. *et al.* Sustainable data analysis with Snakemake. Preprint at https://doi.org/10.12688/f1000research.29032.2 (2021).

12. Privé, F., Arbel, J. & Vilhjálmsson, B. J. LDpred2: better, faster, stronger. *Bioinformatics* **36**, 5424–5431 (2020).
13. R Core Team. *R: A Language and Environment for Statistical Computing*. (R Foundation for Statistical Computing, 2020).
14. Sørensen, E. *et al.* Data Resource Profile: The Copenhagen Hospital Biobank (CHB). *Int. J. Epidemiol.* **50**, 719–720e (2021).
15. Altshuler, D. M. *et al.* Integrating common and rare genetic variation in diverse human populations. *Nature* **467**, 52–58 (2010).
16. Nielsen, J. B. *et al.* Biobank-driven genomic discovery yields new insight into atrial fibrillation biology. *Nat. Genet.* **50**, 1234–1239 (2018).
17. Mahajan, A. *et al.* Fine-mapping type 2 diabetes loci to single-variant resolution using high-density imputation and islet-specific epigenome maps. *Nat. Genet.* **50**, 1505–1513 (2018).
18. Wuttke, M. *et al.* A catalog of genetic loci associated with kidney function from analyses of a million individuals. *Nat. Genet.* **51**, 957–972 (2019).
19. Surakka, I. *et al.* The impact of low-frequency and rare variants on lipid levels. *Nat. Genet.* **47**, 589–597 (2015).
20. Shah, S. *et al.* Genome-wide association and Mendelian randomisation analysis provide insights into the pathogenesis of heart failure. *Nat. Commun.* **11**, 163 (2020).
21. Malik, R. *et al.* Multiancestry genome-wide association study of 520,000 subjects identifies 32 loci associated with stroke and stroke subtypes. *Nat. Genet.* **50**, 524–537 (2018).
22. Jiang, L., Zheng, Z., Fang, H. & Yang, J. A generalized linear mixed model association tool for biobank-scale data. *Nat. Genet.* **53**, 1616–1621 (2021).
23. van der Harst, P. & Verweij, N. Identification of 64 Novel Genetic Loci Provides an Expanded View on the Genetic Architecture of Coronary Artery Disease. *Circ. Res.* **122**, 433–443 (2018).

24. Hoffmann, T. J. *et al.* Genome-wide association analyses using electronic health records identify new loci influencing blood pressure variation. *Nat. Genet.* **49**, 54–64 (2017).

25. Anstee, Q. M. *et al.* Genome-wide association study of non-alcoholic fatty liver and steatohepatitis in a histologically characterised cohort☆. *J. Hepatol.* **73**, 505–515 (2020).