## RESEARCH





Marc Delord<sup>1\*</sup> and Abdel Douiri<sup>1</sup>

## Abstract

Multimorbidity is characterized by the accrual of two or more long-term conditions (LTCs) in an individual. This state of health is increasingly prevalent and poses public health challenges. Adapting approaches to effectively analyse electronic health records is needed to better understand multimorbidity. We propose a novel unsupervised clustering approach to multiple time-to-event health records denoted as multiple state clustering analysis (MSCA). In MSCA, patients' pairwise dissimilarities are computed using patients' state matrices which are composed of multiple censored time-to-event indicators reflecting patients' health history. The use of state matrices enables the analysis of an arbitrary number of LTCs without reducing patients' health trajectories to a particular sequence of events. MSCA was applied to analyse multimorbidity associated with myocardial infarction using electronic health records of 26 LTCs, including conventional cardiovascular risk factors (CVRFs) such as diabetes and hypertension, collected from south London general practices between 2005 and 2021 in 5087 patients using the MSCA R library. We identified a typology of 11 clusters, characterised by age at onset of myocardial infarction, sequences of conventional CVRFs and non-conventional risk factors including physical and mental health conditions. Interestingly, multivariate analysis revealed that clusters were also associated with various combinations of socio-demographic characteristics including gender and ethnicity. By identifying meaningful sequences of LTCs associated with myocardial infarction and distinct socio-demographic characteristics, MSCA proves to be an effective approach to the analysis of electronic health records, with the potential to enhance our understanding of multimorbidity for improved prevention and management.

**Keywords** Electronic health records, Multiple state analysis, Jaccard dissimilarity index, Ward clustering, Multimorbidity, Myocardial infarction

\*Correspondence: Marc Delord marc.delord@kcl.ac.uk <sup>1</sup> School of Life Course & Population Sciences, Department of Population Health Sciences, King's College London, London, UK



© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

## Background

The notion that life-course health trajectories are influenced by early life events and the ever-changing historical context [1], or by later lifestyle associated exposure [2] have dominated the epidemiological field so far.

Accordingly, life-course epidemiology has developed the environmental model-based approach to chronic diseases, focusing on lifestyles and previous exposures that may explain future health conditions. Conceptual life-course models include, for instance, the critical period model [3], the accumulation risk model [4] or the chain risk model [5]. These models emphasize various aspects of life experiences such as the impact of early-life exposures on later health outcomes, the cumulative effects of multiple risk factors and exposure over time, and the interdependence of health events respectively. Conversely, we can assume that the main combinations of early life events and specific socio-demographic contexts result in distinct patterns of life-course health trajectories characterised by specific patterns of long-term conditions (LTCs).

Progress in information technologies and the availability to academic research of publicly funded health databases including routinely collected general practice records represent an opportunity to better evaluate this hypothesis [6]. Electronic health records are digital versions of patients' medical history, containing healthcare-related data such as diagnoses and treatments. The widespread adoption of electronic health records in healthcare systems had led to the accumulation of vast amounts of longitudinal data, which can be harnessed to analyse health outcomes throughout an individual's lifetime.

However, analyzing health records can be challenging. In this setting, the use of unsupervised exploratory analysis may be required to facilitate description of the main patterns of multimorbidity, guiding further research towards relevant underlying conceptual models.

Importantly, the analysis of electronic health records enables data interpretation within patients' socio-demographic contexts, thereby supporting the development of public health preventive strategies targeting potential modifiable risk factors in identified subpopulations.

We propose in this paper a novel unsupervised clustering approach to multiple time-to-event records, denoted as *multiple state analysis*. In the setting of observational studies for instance, this method allows to handle, for each analysed patient a potentially large numbers of LTCs in order to obtain clusters of health trajectories characterised by major sequences of analysed LTCs.

We applied multiple state analysis to electronic health records of LTCs associated with myocardial infarction,

including the conventional cardiovascular risk factors (CVRFs) such as diabetes and hypertension, collected from south London general practices between April 2005 and April 2021. In addition to well-documented trajectories, such as the association of hypertension and diabetes with cardiovascular diseases [7], this research aims at identifying less common patterns of multimorbidity characterized by non-conventional sequences of LTCs.

Myocardial infarction, leading to "heart attack", is one of the leading causes of death in high-income countries [8]. Myocardial infarction is caused by decreased or complete cessation of blood flow in the myocardium and results in irreversible damage to the heart muscle [9]. Most of the time, myocardial infarction is due to underlying coronary artery disease [10]. Conventional modifiable risk factors associated with coronary artery disease and myocardial infarction include smoking, abnormal blood lipid profile, hypertension, diabetes, abdominal obesity, psycho-social factors, diet, physical inactivity and alcohol consumption (protective) [11, 12]. Some non-modifiable risk factors associated with myocardial infarction include advanced age, male gender (males tend to have myocardial infarction earlier in life) and genetics [12, 13].

This paper is organized as follows: in the Background section, multiple state analysis is presented. In the Methods section, an application of multiple state analysis to the analysis of electronic health records of 26 LTCs in patients with myocardial infarction is conducted. Finally in the Multiple state analysis of multimorbidity associated with myocardial infarction section, the proposed method is briefly discussed in light of the study results.

#### Methods

Multiple state analysis is an unsupervised clustering approach to multiple time-to-event records. It is designed to handle and analyse series of events associated with each instance of a population, such as a set of LTCs recorded in a cohort of patients. In the epidemiological setting, the objective is to create a typology of the main patterns of multimorbidity, allowing therefore a simplified description of analysed cohorts in terms of the main underlying patterns of multimorbidity. This objective is achieved by applying the principles of unsupervised clustering analyses [14]: given a relevant metrics, pairwise patients' dissimilarities are computed before a clustering method is applied and a typology of the main health trajectories is defined. Further investigations includes evaluating the association between identified clusters and socio-demographic characteristics and other CVRFs.

#### Individual patient's state matrix and group summaries

In order to compute patients' pairwise dissimilarities, multiple state analysis requires patients' health records to be formatted into multiple indicator tracks, stacked in  $t \times k$  state matrices, t and k being the maximum age observed in the cohort and the total number of recorded LTCs, respectively.

If onset of disease *l* for patient *j* is  $t_{l,j}^{o}$ , this patient's state matrix is noted:  $M_{l}^{i} = I_{l(z \in t_{i,j}^{o})}$  where  $M_{.l}^{j}$  stands for column *l* of matrix  $M^{j}$ , and I is the positive integer indicator function. For instance, in the example here after, examination of patient *j*'s state matrix  $(M^{j})$  indicate that the analysis is conducted on LTCs *a*, *b*, *c*, *d* and *e* from 0 to 99 years, onset of diseases *a*, *b*, *c*, and *e* being 2, 46, 47 and 48 respectively. Patients may also be censored or experience a competing event such as death or may be lost of follow up.

 $\tau$  being patients' censoring times vector,  $c^j = I_{[t \ge \tau_j]}$  represents patient's *j* censoring indicator and  $\bar{c}^j = I_{[t < \tau_j]}$ , the follow up period indicator for patient *j*. In the proposed example, patient *j* is censored at age 83.

Alternatively, in cases EHRs shows evidence that patient *j* is cured from long term condition *l* at time  $t_{l,j}^c < \tau_j$ , their state matrix would be modified such that  $M_{J}^i = I_{[t_{l,j}^o \leq t \leq t_j]}$ . This later expression of *M* is a more general formula of a patient state matrix. From the previous example, if EHRs indicate that patient *j* is cured from diseases *b* at 48 years, for instance, their state matrix  $M'_j$  would be:

		a	b	c	d	e		
	0	(0)	0	0	0	0\		
	1	0	0	0	0	0		
	2	1	0	0	0	0		
	:		÷	÷	÷	:		
	45	1	0	0	0	0		(2)
$\mathbf{M}^{'\mathbf{j}} =$	46	1	1	0	0	0		
	47	1	1	1	0	0	(	
	48	1	0	1	0	1		
		:	:	-	:	:		
	81	1	0	1	0	1		
	82	1	0	1	0	1		
	83	1	0	1	0	1		
	-	:	÷		÷	:		
	99	$\setminus 1$	0	1	0	1/		

## **Dissimilarity index matrix**

The objective pursued when creating a typology of patients based on their profiles of LTCs is to categorise patients such that patients belonging to the same clusters share similar trajectories relatively to patients belonging to other clusters. This process implies therefore the use of a dissimilarity (or distance) metric in order to obtain pairwise patients' profile dissimilarities. Based on the resulting distance matrix or *dissimilarity index matrix*, a clustering method can be applied to finally define a typology.

As multiple state analysis deals with state indicators, the relationship between two patients, regarding a given LTC can be summarised using a  $2 \times 2$  contingency table such as: [(p, r), (s, q)],

- (i) *q* being the number of matching time units during which both patients were affected by a given LTC,
- (ii) *p* being the number of matching time units during which both patients were free from the considered LTC,
- (iii) *s* and *r* being the number of matching time units during which both patients were in different states of health regarding the specified LTC and,
- (iv) *t* being the length of the sequence.

In this setting, the dissimilarity between profiles  $x^i$  and  $x^j$  can be written as:

$$\begin{cases} d(x^i, x^j) = \frac{r+s}{q+r+s+p}, \\ q+r+s+p = t \end{cases}$$

Considering states of illness as more informative than healthy states and omitting p, the number of negative

matches in the denominator, leads to the Jaccard dissimilarity index [15]:

$$d(x^{i}, x^{j}) = \frac{r+s}{q+r+s} = 1 - \frac{q}{q+r+s}$$

Since r + s = t - p - q, we can rewrite the above relation as follows:

$$d(x^{i}, x^{j}) = 1 - \frac{q}{t-p}.$$
(3)

A composite analogue to the Jaccard dissimilarity index when considering indicators from multiple states can be derived as:

$$\begin{cases} d(M^{i}, M^{j}) = 1 - \frac{Q}{t^{*} - P}, \\ Q = \sum_{l=1}^{k} q_{l}, \\ P = \sum_{l=1}^{k} p_{l}, \\ t^{*} = kt. \end{cases}$$
(4)

i.e. *Q* and *P* are the sum of *q* and *p* over all considered LTCs.

Using the matrix notations (1 or 2) from Individual patient's state matrix and group summaries section and setting  $C^i$  as the  $t \times t$  diagonal matrix with  $\bar{c^i}$  as diagonal entries, the censored quantities defined above can be conveniently computed as:

$$Q = \operatorname{tr}\left(\left(\mathbf{M}^{i\prime}\bar{\mathbf{C}}^{i}\bar{\mathbf{C}}^{j}\right)\left(\mathbf{M}^{j\prime}\bar{\mathbf{C}}^{i}\bar{\mathbf{C}}^{j}\right)'\right)$$
$$P = \operatorname{tr}\left(\left(\left(\mathbf{M}^{i}-\mathbb{1}\right)'\bar{\mathbf{C}}^{i}\bar{\mathbf{C}}^{j}\right)\left(\left(\mathbf{M}^{j}-\mathbb{1}\right)'\bar{\mathbf{C}}^{i}\bar{\mathbf{C}}^{j}\right)'\right)$$
$$t^{*} = \operatorname{tr}\left(\left(\mathbb{1}'\bar{\mathbf{C}}^{i}\bar{\mathbf{C}}^{j}\right)\left(\mathbb{1}'\bar{\mathbf{C}}^{i}\bar{\mathbf{C}}^{j}\right)'\right),$$

where tr denotes the trace operator and  $\mathbb{1}$  is the  $t \times k$  matrix with all entries are set to 1.

As an illustration, let's consider health records of two patients, patient *i* having a record of diabetes (*dm*) at 55 and a record of hypertension (*hyp*) at 53, and patient *j* having a record of diabetes at 52 and a record of hypertension at 59.

The state matrices of these patients are:

$$\mathbf{m}^{i} = \begin{array}{cccc} dm & hyp & dm & hyp \\ 50 & 0 & 0 \\ 51 & 0 & 0 \\ 52 & 0 & 0 \\ 53 & 0 & 1 \\ 55 & 0 & 1 \\ 55 & 1 & 1 \\ 55 & 1 & 1 \\ \vdots & \vdots \\ 58 & 1 & 1 \\ 59 & 1 & 1 \\ 60 & 1 & 1 \end{array} \right), \ \mathbf{m}^{j} = \begin{array}{ccc} 50 & 0 & 0 \\ 51 & 0 & 0 \\ 51 & 0 \\ 52 & 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ \vdots & \vdots \\ 1 & 0 \\ 1 & 0 \\ 1 & 1 \end{array} \right)$$

The *composite* Jaccard dissimilarity index proposed in (4) computed between patient *i* and *j* for diabetes and hypertension is computed as follow:

• compute the total number of time units considered:

 $t^* = 11 \times 2 = 22$ ,

• compute the number of matching time units during which both patients were in the same state of health:

$$Q = q_{dm} + q_{hyp} = 6 + 2 = 8$$

• compute the number of matching time units during which both patients were free from the considered LTCs:

$$P = p_{dm} + p_{hyp} = 2 + 3 = 5,$$

• get the Jaccard dissimilarity index using (3):

$$d(m^{i}, m^{j}) = 1 - \frac{Q}{t^{*} - P} = 1 - \frac{8}{22 - 5} = \frac{9}{17}$$

### **Clustering method**

Although multiple state analysis is not restricted to a specific clustering procedure, we have used in this paper the Ward's hierarchical clustering method [16]. At the starting point of this procedure, each instance is considered as a cluster of its own, then clusters are recursively merged such that the resulting cluster structure presents the minimum cost in terms of the *within-clusters* dissimilarity, which often results in compact and well-defined clusters [17]. Although originally proposed in a Euclidean setting, the Ward's method can be generalized to dissimilarity measures, such as the Jaccard dissimilarity, using the Lance-Williams formula and the coefficients associated with Ward's method [18], provided the dissimilarity measure satisfies non-negativity and symmetry.

The Lance-Williams formula is given by the following expression:

$$d(C_i \cup C_j, C_k) = \alpha_i d(C_i, C_k) + \alpha_j d(C_j, C_k) + \beta d(C_i, C_j) + \gamma |d(C_i, C_k) - d(C_j, C_k)|,$$

in which the dissimilarity between merged clusters *i* and *j* and cluster *k* ( $d(C_i \cup C_j, C_k)$ ), is expressed as a linear combination of the dissimilarities between involved clusters without explicit reference to the type of metrics used to compute the dissimilarity matrix between instances. The Lance-Williams coefficients associated with the Ward's method are expressed as follows:

$$\begin{cases} \alpha_i = \frac{|C_i| + |C_k|}{|C_i| + |C_j| + |C_k|} \\ \alpha_j = \frac{|C_j| + |C_k|}{|C_i| + |C_k|} \\ \beta = -\frac{|C_k|}{|C_i| + |C_j| + |C_k|} \\ \gamma = 0 \end{cases}$$

where  $|C_i|$  represent the size of cluster *i*.

## MSCA algorithm time complexity

A critical point inherently associated with MSCA is its computational cost. Instead of clustering a reduced set of long term conditions based on a set of patients, MSCA is a patient-oriented method that produces a typology of patients sharing similar health trajectories. On the one hand, although this strategy allows for capturing patients pairwise longitudinal dissimilarities through the computation of state matrices, the computation cost associated with Ward's clustering method (theoretical time complexity:  $O(n^3)$ ) and the patientwise dissimilarity matrix (time complexity:  $O(n^2)$ ) represents a significant consideration when applying MSCA.

Practically, the use of the *fastclust* C++-implemented routines [19] to perform the Ward's clustering method, through its interfaces to R's *hclust* or Python's *scipy.cluster.hierarchy.linkage*, reduces the observed runtime to  $O(n^2)$  in many cases [20], making it scalable for larger datasets.

For *n* patients, given a symmetric dissimilarity measure, the pairwise dissimilarities between all patients can be stored in an upper triangular matrix of size *n*, corresponding to  $\frac{n(n-1)}{2}$  distinct elements (hence its  $O(n^2)$  time complexity). The time complexity associated with the computation of the Jaccard dissimilarity between two time-aligned sequences of length *t* scales linearly with *t*, i.e. O(t). Therefore, the time complexity associated with the computation of the dissimilarity matrix is  $O(n^2 \cdot l \cdot t)$ , *l* being the number of long-term condition considered. This can be approximated as  $O(n^2)$  when  $l \cdot t$  is much smaller than the number of patients *n*. In scenarios where *n* is large, it may therefore be relevant to optimize the time granularity to effectively reduce MSCA computational cost.

In the next section, we propose an application of MSCA to the analysis of multimorbidity associated with myocardial infarction.

# Multiple state analysis of multimorbidity associated with myocardial infarction

To illustrate *multiple state analysis* we have conducted an analysis of multimorbidity associated with myocardial infarction using electronic health records of 26 long-term conditions including conventional cardiovascular risk factors such as diabetes and hypertension, collected from south London general practices between 2005 and 2021 in 5087 patients.

## Patients and method Primary care registry

We considered electronic health records of 27 common LTCs (including myocardial infarction) in adult patients aged over 18 and registered in 41 general practices in south London between April 2005 and April 2021. Recorded LTCs are listed in Table 1. Briefly, the proposed list includes conventional CVRFs such as hypertension, and diabetes, but also LTCs a priori less related or non-directly related to myocardial infarction such as cancers, chronic kidney disease, asthma, and chronic obstructive pulmonary disease (COPD). Patients' electronic files included the date at which any of the considered LTCs were first ever recorded.

Socio-demographic variables and risk factors collected were: age, gender, ethnicity (asian, black, mixed and other, and white), polymediaction status (defined as eight or more different medications in different BNF (British National Formulary) chapters and sub headings, quintile of locally calculated index of multiple deprivation (IMD) 2019, hypercholesterolaemia (total cholesterol over 5.0 mmol/L) and current or ex smoking habits. Data were provided by the Lambeth DataNet and approval for the analysis of fully anonymised data was granted by Lambeth DataNet Clinical Commissioning Group and Information Governance Steering Group.

## Statistical analysis

The analysis was conducted on patients with a record of myocardial infarction, according to the 3 steps described in Methods section: i) arrange patients individual records into multiple time-to-event indicators stacked in individual patients' state matrices and censoring indicators, ii) compute pairwise patients' dissimilarities on individual state matrices (and censoring indicators) and apply a clustering method, and iii) define a typology.

State matrices were computed considering records associated with the 27 LTCs displayed in Table 1. Figure 1 displays state matrices computed using records of 8 LTCs associated with nine patients randomly sampled from the analysed cohort.

Pairwise dissimilarities between patients were computed using the Jaccard dissimilarity index [15] as described in Dissimilarity index matrix section. The Jaccard coefficient, is a measure of dissimilarity between binary samples, such as state indicators. Finally, agglomerative hierarchical clustering was computed using the Lance-Williams formula with coefficients 

 Table 1
 Long term conditions analysed: long-term conditions are classified according to the 10<sup>th</sup> international classification of diseases (ICD-10)

Category	ICD-10 code	Long-term condition
Neoplasms	C00-C97	Malignant neoplasms
Diseases of the circulatory system	160-164	Stroke
	150	Heart failure
	173	Peripheral vascular disease
	148	Atrial fibrillation
	110	Hypertension
	121	Acute myocardial infarction (MI)
	G45	Transient cerebral ischemic attacks
Certain infectious and parasitic diseases	B15-B19	Viral hepatitis
Diseases of the musculoskeletal system and connective tissue	M05-M06	Rheumatoid arthritis
	M15-M19	Osteoarthritis
	M80-M82	Osteoporosis
Diseases of the digestive system	K50-K52	Inflammatory bowel disease (IBD)
	K70-K76	Chronic liver disease and cirrhosis
Diseases of the genitourinary system	N18	Chronic kidney disease stage 3 to stage 5 (CKD 3–5)
Mental, Behavioral and Neurodevelopmental disorders	F40-F41	Anxiety disorders
	F32	Depression
	F20-F29	Other serious mental illness
	F10	Alcohol dependence
	F11-F19	Other substance dependency
Endocrine, nutritional and metabolic diseases	E10-E14	Diabetes
	E66	Morbid obesity
Diseases of the nervous system	G20	Parkinson's
	G40-G41	Epilepsy
	F00-F03, G30	Dementia
Diseases of the respiratory system	J45-J46	Asthma
	J44	Chronic Obstructive Pulmonary Disease (COPD)

corresponding to the Ward's method, and the point biserial correlation was used to determine the optimal size for the typology given a convenient and workable range (from two to 14) [21].

#### **Graphical representation**

For a better understanding the link between onset of myocardial infarction and associated LTCs in defined clusters, we used a graphical representation displaying health conditions as dots whose x-coordinate represents the median age at onset. On the y axis, dots are conveniently assigned to layers, such that significant transitions are represented by edges oriented from higher to lower levels [22].

## Statistical analysis of cluster trajectories

Socio-demographics, conventional CVRFs and LTC indicators, were displayed as frequencies and percentage or median and interquartile range as appropriate.

Associations between variables and clusters were tested using the Fisher exact test or using Kruskal-Wallis test for numeric variables (Table 2). Multivariate associations between clusters and socio-demographic variables, conventional CVRFs, and LTCs were estimated using logistic regressions where cluster indicators were explained by tested variables. Results were displayed using heatmaps where estimates associated with P values greater than 5% were omitted (Fig. 3). All computations were performed using the R language and environment for statistical computing (version 4.3.0 (2023-04-21)) [23]. State matrices and patients pairwise distances were computed using the MSCA R library [24].

## Results

The study workflow is displayed Fig. 2. Of 856,342 registered patients 5087 (0.59%) had a record of myocardial infarction (Tables 2 and 3). This corresponds to an incidence rate of 100.95 cases per 100,000 person-years. The



## Patients' state matrices

**Fig. 1** Examples of patients' state matrices: censored status of 8 long-term conditions are considered from 0 to 104 years old in nine patient randomly sampled from patients' cohort. Patients' status are represented by state indicators for each displayed long-term condition. If during its follow up a patient remains free from a given long-term conditions, the corresponding state indicator will remain zeros from the age of zero to the age at which patient's follow up ends. If for instance, a patient is diagnosed with hypertension at 50 years, patient's state indicator for hypertension will be zero from zero to 49 years old, and one from 50 years old to the end of patient's follow up

median age at onset of myocardial infarction was 61.5. Among patients with a record of myocardial infarction, 32.2% were female, 49.3% were white and the majority (66.7%) of the population belonged to IMD quintile 1–2 (most deprived vs quintile 3–5 (less deprived)). The main LTCs associated with myocardial infarction were hypertension (67.2%) and diabetes (39.2%) and the median number of LTCs associated with myocardial infarction was 5 (IQR: 3). The median follow-up was 9.6 years, 61.7% of patients were censored or died at end of follow-up (39.3%) (Table 2).

After pairwise patients' dissimilarity computation and clustering, an 11 clusters partition was retained. This partition size corresponds to a local maximum of the point biserial correlation for a range extending from two to 14 clusters [21].

#### Typology annotation

Clusters were ordered by decreasing frequency, and numbered from 1 to 11. Cluster 1 represents 1017 patients (20.0%) and clusters 11 represents 172 patients (3.4%). Table 2 displays socio-demographic variables and risk factors distribution across clusters. Alternatively, Fig. 3 displays the multivariate log-odds ratio of sociodemographic variables and risk factors (upper panel) and LTCs (lower panel) across clusters. Additionally a graphical representations of time at onset of myocardial infarction and highly associated LTCs in proposed in Fig. 4, and Fig. 5 displays an interpretation of the proposed typology according to clusters' prevalence in hypertension and diabetes. Finally, Table 4 displays the number of sequences of length 2 across the different clusters and detail sequences shared by at least 30% of patients

Table 2 Distributio	n of socio-demograp	hic variables	and risk fa	actors in my	ocardial in	farction pa	tients acro	ss defined	clusters					
Cluster number		-	2	m	4	5	9	2	8	6	10	11	Total	þ
Gender	Female	371 (36.5)	129 (20.3)	165 (26.0)	158 (29.8)	281 (54.7)	87 (23.8)	66 (19.3)	133 (41.2)	99 (31.5)	97 (41.1)	54 (31.4)	1640 (32.2)	<0.001
Age at onset	Median (IQR <sup>a</sup> )	68.1 (16.9)	45.7 (10.8)	56.2 (14.7)	63.2 (11.4)	80.6 (12.1)	54.6 (10.5)	53.1 (6.3)	69.1 (14.8)	63.1 (8.0)	72.2 (9.5)	49.4 (10.4)	61.5 (20.1)	<0.001
Ethnicity	White	511 (50.2)	284 (44.6)	393 (61.9)	242 (45.6)	268 (52.1)	114 (31.1)	184 (53.8)	165 (51.1)	181 (57.6)	82 (34.7)	82 (47.7)	2506 (49.3)	<0.001
	Asian	81 (8.0)	59 (9.3)	32 (5.0)	46 (8.7)	24 (4.7)	66 (18.0)	30 (8.8)	30 (9.3)	18 (5.7)	42 (17.8)	11 (6.4)	439 (8.6)	
	Black	114 (11.2)	107 (16.8)	41 (6.5)	78 (14.7)	37 (7.2)	71 (19.4)	16 (4.7)	47 (14.6)	11 (3.5)	40 (16.9)	30 (17.4)	592 (11.6)	
	Other/mixed/missing	311 (30.6)	187 (29.4)	169 (26.6)	165 (31.1)	185 (36.0)	115 (31.4)	112 (32.7)	81 (25.1)	104 (33.1)	72 (30.5)	49 (28.5)	1550 (30.5)	
IMD <sup>b</sup> quintile	>2 - Least deprived	316 (31.3)	211 (33.3)	192 (30.6)	197 (37.5)	167 (32.6)	114 (31.5)	136 (40.0)	112 (35.0)	118 (38.2)	77 (32.6)	43 (25.4)	1683 (33.3)	0.007
	<3 - Most deprived	695 (68.7)	422 (66.7)	436 (69.4)	329 (62.5)	346 (67.4)	248 (68.5)	204 (60.0)	208 (65.0)	191 (61.8)	159 (67.4)	126 (74.6)	3364 (66.7)	
Hypercholesterol- aemia <sup>c</sup>	Yes	758 (76.2)	459 (73.0)	465 (74.4)	406 (76.5)	349 (70.8)	278 (77.2)	238 (72.8)	242 (76.8)	212 (70.7)	166 (73.5)	126 (74.6)	3699 (74.4)	
Smoking ever	Yes	770 (77.4)	497 (79.0)	540 (86.4)	370 (69.7)	288 (58.4)	275 (76.4)	267 (81.7)	221 (70.2)	222 (74.0)	148 (65.5)	144 (85.2)	3742 (75.3)	
Polymedication <sup>d</sup>	Yes	434 (43.3)	179 (28.7)	207 (33.3)	174 (33.3)	211 (42.5)	147 (40.7)	100 (30.0)	126 (39.4)	109 (36.5)	117 (51.5)	60 (34.9)	1864 (37.4)	
Number of LTC <sup>e</sup>	Median (IQR)	6.0 (3.0)	5.0 (3.0)	6.0 (4.0)	5.0 (4.0)	5.0 (2.8)	5.0 (4.0)	4.0 (3.0)	6.0 (2.0)	5.0 (3.0)	6.0 (4.0)	6.0 (5.0)	5.0 (3.0)	<0.001
Status	Death	454 (44.6)	151 (23.7)	171 (26.9)	193 (36.3)	346 (67.3)	115 (31.4)	134 (39.2)	122 (37.8)	163 (51.9)	113 (47.9)	35 (20.3)	1997 (39.3)	<0.001
Total N (%)		1017 (20.0)	637 (12.5)	635 (12.5)	531 (10.4)	514 (10.1)	366 (7.2)	342 (6.7)	323 (6.3)	314 (6.2)	236 (4.6)	172 (3.4)	5087	
<sup>a</sup> Interquartile range <sup>b</sup> Index of multiple dep	ivation (auintile 1 (most de	eprived)—2 vs 3	1-5 (less depr	ived), 2019)										

-2 Ľ 5

<sup>c</sup> Total cholesterol over 5.0 mmol/L

<sup>d</sup> Eight or more different medication in different chapters of the British National Formulary (BNF) and sub-headings

<sup>e</sup> Long-term condition



Fig. 2 Study workflow

in the corresponding cluster as well as the percentage of patients following this sequence in that cluster.

#### **Clusters** examination

Clusters showed distinct demographic characteristics and levels of the conventional CVRFs such as age at onset (clusters 5 and 10), hypertension (cluster 2, 4, 5) and diabetes (clusters 6 and 10). Conversely, other clusters were highly associated with specific non-conventional CVRFs, such as higher number of recorder LTCs (clusters 1, 3 and 11), mental health conditions (cluster 3), Osteoarthritis (cluster 8) and asthma (cluster 11). Finally, other clusters (7 and 9) were characterised with significantly lower prevalence of hypertension and/or diabetes, a lower number of recorded LTCs as well as lower prevalence of mental health conditions. Non-conventional CVRFs associated with these clusters were peripheral vascular disease and COPD (cluster 7) and atrial fibrillation (cluster 9) (Tables 2, 3 and Fig. 5).

## Conclusions

We have presented in this paper a novel approach to the unsupervised analysis of multiple time-to-event records denoted as *multiple state analysis* with application to the analysis of electronic health records in patients with myocardial infarction.

Similar to clustering methodology developed in other fields of research [25–27], *multiple state analysis* follows the general principle of cluster analysis where individuals are grouped according to a systematic rule. However, unlike the analysis of sequences developed in the field of the social sciences, in which multiple aspects of the social experience, such as marital and employment status, are jointly analysed [26, 27], the proposed method does not deal primarily with constructed sequence of data but binary coding of individual records into states matrices and associated follow up indicators. Importantly, if meaningful sequences of LTCs may be derived from resulting clusters, electronic health records and associated state matrices are not sequences of events.

This aspect is important in the context of epidemiological research where patients' state of health are often characterized by the accrual of two or more LTCs, that may occur simultaneously or in a short period of time [28]. In this setting, the simple record of events and the subsequent analysis of resulting sequences may be misleading and fail to fully render the evolution of patients'

Cluster number	-	2	m	4	Ω.	9	7	œ	6	10	11	Total
Cancer	316 (31.1)	65 (10.2)	94 (14.8)	72 (13.6)	84 (16.3)	29 (7.9)	53 (15.5)	44 (13.6)	60 (19.1)	41 (17.4)	22 (12.8)	880 (17.3)
Stroke	261 (25.7)	100 (15.7)	88 (13.9)	71 (13.4)	87 (16.9)	61 (16.7)	33 (9.6)	59 (18.3)	33 (10.5)	35 (14.8)	26 (15.1)	854 (16.8)
Heart Failure	314 (30.9)	158 (24.8)	120 (18.9)	152 (28.6)	150 (29.2)	114 (31.1)	111 (32.5)	95 (29.4)	114 (36.3)	77 (32.6)	50 (29.1)	1455 (28.6)
Peripheral Vascular Disease	144 (14.2)	72 (11.3)	65 (10.2)	58 (10.9)	37 (7.2)	74 (20.2)	50 (14.6)	40 (12.4)	25 (8.0)	34 (14.4)	18 (10.5)	617 (12.1)
Atrial Fibrillation	201 (19.8)	70 (11.0)	77 (12.1)	119 (22.4)	134 (26.1)	44 (12.0)	72 (21.1)	77 (23.8)	83 (26.4)	42 (17.8)	17 (9.9)	936 (18.4)
Hypertension	657 (64.6)	510 (80.1)	317 (49.9)	523 (98.5)	396 (77.0)	262 (71.6)	133 (38.9)	222 (68.7)	139 (44.3)	169 (71.6)	88 (51.2)	3416 (67.2)
Transient Ischemic Attack	73 (7.2)	40 (6.3)	25 (3.9)	67 (12.6)	37 (7.2)	25 (6.8)	19 (5.6)	23 (7.1)	19 (6.1)	14 (5.9)	9 (5.2)	351 (6.9)
Viral Hepatitis (B & C)	34 (3.3)	11 (1.7)	9 (1.4)	2 (0.4)	0 (0:0)	4 (1.1)	1 (0.3)	3 (0.9)	0(0.0) 0	5 (2.1)	3 (1.7)	72 (1.4)
Rheumatoid Arthritis	58 (5.7)	7 (1.1)	14 (2.2)	8 (1.5)	12 (2.3)	7 (1.9)	3 (0.9)	9 (2.8)	7 (2.2)	2 (0.8)	5 (2.9)	132 (2.6)
IBD <sup>a</sup>	43 (4.2)	8 (1.3)	6 (0.9)	6 (1.1)	3 (0.6)	1 (0.3)	4 (1.2)	6 (1.9)	4 (1.3)	1 (0.4)	8 (4.7)	90 (1.8)
CKD <sup>b</sup> 3–5	303 (29.8)	137 (21.5)	104 (16.4)	178 (33.5)	198 (38.5)	115 (31.4)	67 (19.6)	89 (27.6)	98 (31.2)	99 (41.9)	32 (18.6)	1420 (27.9)
Liver Disease	16 (1.6)	6 (0.9)	30 (4.7)	7 (1.3)	3 (0.6)	9 (2.5)	0 (0:0)	3 (0.9)	4 (1.3)	4 (1.7)	5 (2.9)	87 (1.7)
Anxiety Disorders	213 (20.9)	155 (24.3)	325 (51.2)	86 (16.2)	65 (12.6)	44 (12.0)	32 (9.4)	49 (15.2)	25 (8.0)	24 (10.2)	65 (37.8)	1083 (21.3)
Depression	288 (28.3)	136 (21.4)	474 (74.6)	69 (13.0)	57 (11.1)	81 (22.1)	38 (11.1)	40 (12.4)	37 (11.8)	33 (14.0)	81 (47.1)	1334 (26.2)
Alcohol Dependence	56 (5.5)	45 (7.1)	218 (34.3)	31 (5.8)	6 (1.2)	14 (3.8)	9 (2.6)	6 (1.9)	4 (1.3)	3 (1.3)	22 (12.8)	414 (8.1)
Serious Mental Illness	83 (8.2)	20 (3.1)	50 (7.9)	11 (2.1)	5 (1.0)	9 (2.5)	2 (0.6)	2 (0.6)	2 (0.6)	1 (0.4)	6 (3.5)	191 (3.8)
Substance Use (Dependency)	59 (5.8)	12 (1.9)	67 (10.6)	4 (0.8)	6 (1.2)	6 (1.6)	4 (1.2)	1 (0.3)	2 (0.6)	1 (0.4)	19 (11.0)	181 (3.6)
Diabetes	314 (30.9)	240 (37.7)	206 (32.4)	131 (24.7)	110 (21.4)	365 (99.7)	124 (36.3)	111 (34.4)	89 (28.3)	235 (99.6)	67 (39.0)	1992 (39.2)
Obesity (morbidly obese)	61 (6.0)	83 (13.0)	36 (5.7)	27 (5.1)	13 (2.5)	46 (12.6)	28 (8.2)	32 (9.9)	11 (3.5)	27 (11.4)	27 (15.7)	391 (7.7)
Osteoarthritis	323 (31.8)	134 (21.0)	140 (22.0)	134 (25.2)	134 (26.1)	76 (20.8)	69 (20.2)	317 (98.1)	81 (25.8)	79 (33.5)	28 (16.3)	1515 (29.8)
Osteoporosis	59 (5.8)	5 (0.8)	24 (3.8)	28 (5.3)	63 (12.3)	6 (1.6)	11 (3.2)	25 (7.7)	23 (7.3)	14 (5.9)	2 (1.2)	260 (5.1)
Parkinson's	27 (2.7)	5 (0.8)	10 (1.6)	9 (1.7)	9 (1.8)	3 (0.8)	5 (1.5)	4 (1.2)	4 (1.3)	6 (2.5)	0 (0:0)	82 (1.6)
Epilepsy	103 (10.1)	23 (3.6)	13 (2.0)	7 (1.3)	1 (0.2)	1 (0.3)	6 (1.8)	4 (1.2)	4 (1.3)	2 (0.8)	6 (3.5)	170 (3.3)
Dementia	134 (13.2)	20 (3.1)	37 (5.8)	50 (9.4)	107 (20.8)	11 (3.0)	33 (9.6)	38 (11.8)	46 (14.6)	27 (11.4)	6 (3.5)	509 (10.0)
Asthma	287 (28.2)	44 (6.9)	77 (12.1)	32 (6.0)	52 (10.1)	26 (7.1)	28 (8.2)	39 (12.1)	19 (6.1)	19 (8.1)	165 (95.9)	788 (15.5)
COPD <sup>c</sup>	265 (26.1)	42 (6.6)	133 (20.9)	66 (12.4)	60 (11.7)	21 (5.7)	86 (25.1)	49 (15.2)	44 (14.0)	25 (10.6)	48 (27.9)	839 (16.5)
Total N (%)	1017 (20.0)	637 (12.5)	635 (12.5)	531 (10.4)	514 (10.1)	366 (7.2)	342 (6.7)	323 (6.3)	314 (6.2)	236 (4.6)	172 (3.4)	5087
<sup>a</sup> Inflammatory Bowel Disease												

Table 3 Distribution of the 26 analyzed comorbidities in myocardial infarction patients across defined clusters

<sup>b</sup> Chronic Kidney Disease stage 3 to stage 5

<sup>c</sup> Chronic Obstructive Pulmonary Disease

	, et	,et?	, et B	ter A	iet S	, et 6	ter 1	ver 8	ier of	ter 10	xet 1	
Chi	, Chr	, Chr	, Chr	S. Chr	, Chr	, Chr	S CHI	S. Chr	S CHI	S CIN	3°	
0.37	-1.40	-0.36	0.18	1.68	-0.43	-0.47	0.36	0.10	0.63	-0.77	Age at onset (per 10 year)	
ns	ns	ns	-0.24	0.37	ns	ns	ns	ns	ns	ns	Female gender	
ns	ns	-0.93	ns	ns	1.26	ns	ns	-0.64	1.42	ns	Ethnicity-Asian (ref=white)	
ns	0.67	-0.90	0.36	-0.74	1.04	-1.01	ns	-1.32	0.83	0.46	Ethnicity-Black (ref=white)	
ns	ns	-0.37	ns	ns	0.59	ns	ns	ns	0.35	ns	Ethnicity-Other/mixed/missing (ref=white	te)
ns	ns	ns	ns	ns	ns	-0.30	ns	ns	ns	ns	IMD <sup>a</sup> quintile 1-2 (most deprived, ref=q	uintile 3-5)
ns	ns	-0.21	ns	ns	0.31	ns	ns	ns	ns	ns	Polymedication <sup>b</sup>	
ns	ns	ns	ns	-0.33	0.30	ns	ns	ns	ns	ns	Hypercholesterolaemia (total cholester	ol over 5.0 mmol/L)
0.64	ns	0.65	-0.42	-1.40	ns	-0.80	0.35	-1.08	ns	0.97	Number of LTC <sup>c</sup>	
0.33	-0.39	0.45	ns	ns	ns	ns	ns	ns	ns	ns	Smoking ever	
1.10	-0.61	ns	-0.44	ns	-0.82	ns	-0.45	ns	ns	-0.55	Cancer	
0.73	ns	ns	-0.56	ns	ns	-0.60	ns	-0.60	-0.44	ns	Stroke	
ns	ns	-0.50	ns	ns	ns	ns	ns	0.30	ns	ns	Heart failure	
0.29	ns	-0.36	ns	-0.57	0.48	0.37	ns	ns	ns	ns	Peripheral vascular disease	
ns	-0.56	ns	0.29	0.44	-0.52	ns	ns	0.45	ns	-0.68	Atrial fibrillation	Multiveriete
-0.21	0.92	-0.79	>3	0.58	ns	-1.25	-0.30	-1.09	ns	-0.96	Hypertension	Multivariate
ns	ns	-0.84	0.87	ns	ns	ns	ns	ns	ns	ns	Transient Ischaemic Attack	<b>a</b> 4
1.70	ns	-1.69	ns	ns	ns	ns	ns	ns	ns	ns	Viral hepatitis (B & C)	2
1.22	ns	ns	ns	ns	ns	-1.20	ns	ns	ns	ns	Rheumatoid arthritis	0
1.33	ns	ns	ns	-1.36	ns	ns	ns	ns	ns	1.78	Inflammatory bowel disease (IBD)	2
ns	-0.33	-0.29	ns	0.50	ns	-0.40	-0.36	ns	0.44	ns	CKD 3-5 <sup>d</sup>	4
ns	ns	0.88	ns	ns	ns	ns	ns	ns	ns	ns	Liver disease	— .
-0.21	0.43	0.99	ns	-0.34	-0.43	-0.75	ns	-0.94	ns	0.56	Anxiety disorders	
ns	-0.34	2.57	-0.84	-0.93	ns	-0.85	-0.99	-0.63	-0.64	0.58	Depression	
-1.01	ns	2.69	ns	-1.78	-0.66	-1.20	-1.06	-1.78	-1.54	ns	Alcohol dependence	
1.31	ns	ns	ns	-0.99	ns	-1.55	ns	ns	-2.08	-1.33	Serious Mental Illness	
0.67	-0.68	ns	-1.28	ns	ns	ns	-2.03	ns	ns	ns	Substance use (dependency)	
-0.51	-0.23	ns	-0.92	-1.00	>3	ns	-0.34	-0.43	>3	ns	Diabetes	
ns	0.86	-0.44	ns	-0.98	ns	ns	ns	-0.64	ns	1.06	Obesity (morbid obesity)	
ns	-0.47	-0.46	-0.37	-0.37	-0.54	-0.44	>3	ns	ns	-1.54	Osteoarthritis	
ns	-1.66	ns	ns	0.89	ns	ns	ns	ns	0.71	ns	Osteoporosis	
0.69	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	Parkinson's	
2.07	ns	-1.69	-1.02	-2.98	-2.12	ns	-1.27	ns	ns	ns	Epilepsy	
0.27	-1.22	-0.62	ns	1.01	-1.35	ns	ns	0.52	ns	-1.04	Dementia	
1.02	-0.79	-1.03	-1.07	-0.35	-0.78	-0.90	-0.64	-1.04	-0.68	>3	Asthma	
0.45	-0.87	0.32	ns	ns	-0.95	0.97	ns	ns	ns	-0.71	Chronic obstructive pulmonary disease	e (COPD)

<sup>a</sup>Index of multiple deprivation (quintile 1 (most deprived)-2 vs 3-5 (less deprived), 2019)

<sup>b</sup>Eight or more different medication in different chapters of the British National Formulary (BNF) and sub-headings

<sup>c</sup>Long-term condition

<sup>d</sup>Chronic kidney disease stage 3 to stage 5

**Fig. 3** Log-odds ratio of socio-demographic variables and risk factors associated (upper panel), and acute and long-term conditions (lower panel) across clusters: values are derived from multivariate logistic regressions where clusters' indicators were explained by displayed variables. Positive log-odds ratio illustrate an over representation of the corresponding traits in a given cluster as compared to other analysed patients. Estimates associated with P values greater than 5% were omitted

states of health. This limitation is due to i) the multiplicity of sequences that can be derived from a unique set of electronic health records, ii) the non-univocal interpretation of electronic health records in presence of ties, and iii) the non-mutual exclusivity of events under investigation, such as the onset of multiple LTCs, for instance. One the other hand, the definition and use of state matrices allow to conveniently measure distances between



**Fig. 4** Graphical representation of clusters: long-term conditions are represented by colored dots. Dots x-coordinate represents the median age at onset of the given long-term condition. Dots are presented on the y axis according to layers, such that no significant transitions exist between dots from the same layer. Significant transitions are represented by edges oriented from a higher level to a lower levels

patients without reducing patients' health trajectories to a particular sequence. This allows to handle arbitrary large number of potentially concomitant and non-exclusive LTCs without making a priori choices regarding the importance of input variables, as it is the case in sequence analyses.

Other important aspects of unsupervised clustering methods involve the choice of a dissimilarity metric and a clustering algorithm. We proposed in our study the Jaccard dissimilarity index associated the Ward's hierarchical clustering approach. Also referred to as the binary metric, the Jaccard index allows to conveniently compare multiple instances based on binary attributes such as state indicators. In the clinical setting, the Jaccard index provides also a meaningful epidemiological interpretation: it represents, for two patients and for a given longterm condition, the number of matching time units these patients spent in different states of health over the period either patient have presented a record of the investigated disease. Conversely, the longer two patients remained simultaneously in the same state of health, the smaller the Jaccard index and the more likely these patients to be



Cardiovascular risk factors; "Long-term condition; "Non-White ethnicity d Inflammatory bowel disease; "Chronic obstructive pulmonary disease <sup>1</sup>Chronic kidney disease stage 3 to stage 5

**Fig. 5** Graphical typology annotation: The typology is divided in two main branches. The conventional cardiovascular risk factors branch includes clusters associated with hypertension (clusters 2, 4 and 5) as well as clusters associated with diabetes (clusters 6 and 10). The other main branch is composed by clusters characterised by non conventional risk factors (clusters 1, 3, 7, 8, 9 and 11) among which clusters 7 and 9 present lower prevalence of hypertension and recorded long-term conditions

co-clustered, which meets assigned objectives. Another interesting feature associated with the use of the Jaccard index is a certain robustness to censored measurements. This is due to the implicit censoring of periods simultaneously free from a given condition when computing the index between two patients, much like when one of them is actually censored. The Ward's method used in this paper aim at minimizing the within-cluster dissimilarity and results often in compact and well-defined clusters [17] and represents a consensual choice in clustering analyses [26, 27, 29, 30].

In an application of *multiple state analysis* to multimorbidity associated with myocardial infraction, we analysed 5093 patients with respect to up to 26 LTCs and various follow-up times. Following our strategy, patients were mapped back to a limited number of clusters, characterised by distinct patterns of multimorbidity and sequences of LTCs. Interestingly, resulting clusters had also distinctive socio-demographic characteristics and levels of the conventional CVRFs such as age at onset of myocardial infarction (clusters 5 and 10), hypertension (clusters 2, 4, 5) and diabetes (clusters 6 and 10). Conversely, other clusters were highly associated with specific non-conventional risk factors, such as higher number of recorder LTCs (clusters 1, 3 and 11), mental health conditions (cluster 3), osteoarthritis (cluster 8) and asthma (cluster 11). Of note, 2 clusters (7 and 9) were characterised by significantly lower prevalence of hypertension and/or diabetes, a lower number of recorded LTCs as well as lower prevalence

Cluster#         Number of sequences*         From         To         Support <sup>b</sup> (%)         Confidence <sup>4</sup> (%)           1         904         Myocardial Infarction         Censoring         53.1         53.1           2         735         Myocardial Infarction         Death         41.6         41.6           2         735         Myocardial Infarction         Death         41.6         41.6           3         828         Myocardial Infarction         Censoring         74.7         74.7           4         680         Myocardial Infarction         Hypertension         33.6         33.6           3         828         Myocardial Infarction         Censoring         70.2         70.2           4         680         Hypertension         Myocardial Infarction         61.7         82.7           4         680         Hypertension         Myocardial Infarction         75.5         68.2           5         556         Myocardial Infarction         Death         35.2         35.2           6         623         Diabetes         Myocardial Infarction         88.8         86.0           7         526         Myocardial Infarction         Censoring         67.2         67.2 <th></th> <th></th> <th>Sequences</th> <th></th> <th>Statistics</th> <th></th>			Sequences		Statistics	
1     904     Myocardial Infarction     Censoring     53.1     53.1       1     Hypertension     Myocardial Infarction     44.2     68.5       2     735     Myocardial Infarction     Censoring     74.7     74.7       1     Hypertension     Myocardial Infarction     46.5     58.0       3     828     Myocardial Infarction     Hypertension     33.6     33.6       3     828     Myocardial Infarction     61.7     82.7       2     Anxiety     Myocardial Infarction     77.5     75.5       4     680     Hypertension     Myocardial Infarction     78.3     79.5       4     680     Hypertension     Myocardial Infarction     78.3     79.5       5     556     Myocardial Infarction     Censoring     62.0     62.0       6     623     Diabets     Myocardial Infarction     52.5     68.2       7     526     Myocardial Infarction     Censoring     30.9     30.9       7     526     Myocardial Infarction     Censoring     60.5     60.5       7     526     Myocardial Infarction     Censoring     60.5     60.5       7     526     Myocardial Infarction     Censoring     61.5     61.5	Cluster#	Number of sequences <sup>a</sup>	From	То	Support <sup>b</sup> (%)	Confidence <sup>c</sup> (%)
HypertensionMyocardial Infarction44.268.52735Myocardial InfarctionDeath11.614.62735Myocardial InfarctionMyocardial Infarction45.558.0382.8Myocardial InfarctionCensoring70.270.270.230DepressionMyocardial Infarction61.782.740DepressionMyocardial Infarction78.378.5468.0HypertensionMyocardial Infarction78.379.55Myocardial InfarctionDeath35.262.05Myocardial InfarctionDeath35.262.05Myocardial InfarctionDeath35.262.05Myocardial InfarctionDeath35.262.06Myocardial InfarctionDeath35.262.06Myocardial InfarctionDeath62.062.0752.6Myocardial InfarctionDeath62.06Myocardial InfarctionDeath62.062.0752.6Nyocardial InfarctionDeath35.063.5752.6Myocardial InfarctionDeath36.035.0752.6Myocardial InfarctionDeath36.035.0752.6Myocardial InfarctionDeath36.035.0752.6OsteoarthritisMyocardial Infarction37.037.0852.7OsteoarthritisMyocardial In	1	904	Myocardial Infarction	Censoring	53.1	53.1
1         Myocardial Infarction         Death         41.6         41.6           2         73         Myocardial Infarction         Censoring         74.7         74.7           3         Myocardial Infarction         Hypertension         Myocardial Infarction         33.6         33.6           3         828         Myocardial Infarction         Censoring         70.2         70.2           4         680         Myocardial Infarction         Myocardial Infarction         36.0         73.5           4         680         Myocardial Infarction         Censoring         62.0         62.0           5         Myocardial Infarction         Death         51.5         61.5           5         Myocardial Infarction         Death         51.5         61.5           6         Myocardial Infarction         Death         51.5         61.5           7         Myocardial Infarction         Censoring         62.5         62.2         62.2           6         Opticardial Infarction         Nocardial Infarction         Myocardial Infarction         83.8         80.0           7         S         Diabetes         Myocardial Infarction         Resoring         62.5         60.5           7			Hypertension	Myocardial Infarction	44.2	68.5
2     735     Myocardial Infarction     Censoring     747     747       4     Myocardial Infarction     Myocardial Infarction     65     580       3     828     Myocardial Infarction     Censoring     702     702       4     680     Myocardial Infarction     Censoring     702     702       4     680     Myocardial Infarction     Censoring     702     702       4     680     Myocardial Infarction     Censoring     702     702       5     680     Myocardial Infarction     Censoring     703     795       600     Myocardial Infarction     Death     515     682     752       7     76     Myocardial Infarction     Death     525     682       7     76     Myocardial Infarction     Censoring     702     767       623     Diabetes     Myocardial Infarction     858     860       7     76     Myocardial Infarction     Censoring     615     615       1     Myocardial Infarction     Myocardial Infarction     858     860       7     747     747     Myocardial Infarction     Censoring     615     615       1     Myocardial Infarction     Death     316     316 <td< td=""><td></td><td></td><td>Myocardial Infarction</td><td>Death</td><td>41.6</td><td>41.6</td></td<>			Myocardial Infarction	Death	41.6	41.6
Age and a set of the	2	735	Myocardial Infarction	Censoring	74.7	74.7
3         Myocardial Infarction         Hypertension         33.6         33.6           3         828         Myocardial Infarction         Censoring         70.2         70.2           4         680         Hypertension         Myocardial Infarction         37.6         73.5           4         680         Hypertension         Myocardial Infarction         78.3         79.5           5         556         Myocardial Infarction         Death         32.6         68.2           6         Myocardial Infarction         Death         61.5         61.5           6         Hypertension         Myocardial Infarction         82.6         68.2           6         Myocardial Infarction         Censoring         30.9         30.9           6         Myocardial Infarction         Censoring         67.2         67.2           7         Myocardial Infarction         Censoring         67.2         67.2           7         Myocardial Infarction         Censoring         60.5         60.5           Myocardial Infarction         Censoring         61.5         61.5           8         Myocardial Infarction         Censoring         61.5         61.5           9         Myocardial			Hypertension	Myocardial Infarction	46.5	58.0
3     828     Myocardial Infarction     Censoring     70.2     70.2       4     680     Anxiety     Myocardial Infarction     61.7     62.7       4     680     Myocardial Infarction     78.3     79.5       Myocardial Infarction     Censoring     62.0     62.0       5     56     Myocardial Infarction     Death     51.2     63.2       6     55     Myocardial Infarction     Death     52.5     63.2       6     623     Diabetes     Myocardial Infarction     55.8     63.2       7     526     Diabetes     Myocardial Infarction     61.5     63.2       7     526     Myocardial Infarction     Censoring     60.5     60.5       7     526     Myocardial Infarction     Censoring     61.5     60.5       7     526     Myocardial Infarction     Death     38.3     38.3       8     68.2     Myocardial Infarction     Death     38.3     38.3       8     58.2     Myocardial Infarction     Death     38.3     38.3       9     58.4     Myocardial Infarction     Death     38.3     38.3       9     Myocardial Infarction     Death     38.3     38.3       9     My			Myocardial Infarction	Hypertension	33.6	33.6
PerformantDepressionMyocardial Infarction61.782.7AnxietyMyocardial Infarction37.673.54HypertensionMyocardial Infarction78.379.55Myocardial InfarctionCensoring62.062.0555.6Myocardial InfarctionDeath61.561.56Myocardial InfarctionDeath61.568.26Myocardial InfarctionDeath52.568.26Myocardial InfarctionCensoring30.930.966.3Myocardial InfarctionCensoring67.267.26Myocardial InfarctionCensoring67.267.277.4Myocardial InfarctionDeath30.630.672.6Myocardial InfarctionDeath36.536.575.6Myocardial InfarctionDeath38.338.378Myocardial InfarctionDeath36.558.58Myocardial InfarctionDeath36.558.59514Myocardial InfarctionCensoring54.636.59514Myocardial InfarctionDeath36.536.59514Myocardial InfarctionDeath36.536.59514Myocardial InfarctionDeath36.636.59514Myocardial InfarctionDeath36.536.59514Myocardial InfarctionDeath36.536.5 <td>3</td> <td>828</td> <td>Myocardial Infarction</td> <td>Censoring</td> <td>70.2</td> <td>70.2</td>	3	828	Myocardial Infarction	Censoring	70.2	70.2
AnietyMyocardial Infarction37.673.54680HypertensionMyocardial Infarction78.379.55Myocardial InfarctionDeath62.062.05556Myocardial InfarctionDeath61.561.5HypertensionMyocardial InfarctionDeath62.062.0662.3Myocardial InfarctionDeath61.561.5Myocardial InfarctionDeath61.561.561.56Myocardial InfarctionCensoring30.930.9662.3DiabetesMyocardial Infarction67.267.2Myocardial InfarctionCensoring67.263.763.7752.6Myocardial InfarctionDeath30.630.6752.6Myocardial InfarctionDeath30.630.6752.6Myocardial InfarctionDeath36.630.6752.6Myocardial InfarctionDeath36.558.58Myocardial InfarctionMyocardial Infarction82.483.9951.4Myocardial InfarctionDeath51.636.5951.4Myocardial InfarctionDeath36.436.5951.4Myocardial InfarctionDeath36.436.510Myocardial InfarctionDeath51.636.5951.4Myocardial InfarctionCensoring52.536.511Myocardial InfarctionDeath <td></td> <td></td> <td>Depression</td> <td>Myocardial Infarction</td> <td>61.7</td> <td>82.7</td>			Depression	Myocardial Infarction	61.7	82.7
4680HypertensionMyocardial Infarction78.379.5Myocardial InfarctionCensoring62.062.05HMyocardial InfarctionDeath35.235.25HypertensionMyocardial Infarction52.568.2Myocardial InfarctionCensoring30.930.96623DiabetesMyocardial Infarction85.886.07Agardial InfarctionCensoring67.267.2Myocardial InfarctionMyocardial Infarction85.886.07Myocardial InfarctionCensoring67.267.2Myocardial InfarctionDeath30.630.67S26Myocardial InfarctionCensoring60.560.58Myocardial InfarctionCensoring60.560.58Myocardial InfarctionCensoring82.483.98Myocardial InfarctionMyocardial Infarction82.483.99Myocardial InfarctionMyocardial Infarction82.568.59Myocardial InfarctionDeath36.536.59AffMyocardial InfarctionCensoring45.547.510Myocardial InfarctionDeath16.651.611Myocardial InfarctionCensoring7.27.912Myocardial InfarctionCensoring52.660.513Myocardial InfarctionCensoring50.650.614Myocardial Infarction </td <td></td> <td></td> <td>Anxiety</td> <td>Myocardial Infarction</td> <td>37.6</td> <td>73.5</td>			Anxiety	Myocardial Infarction	37.6	73.5
Myocardial InfarctionCensoring62.062.05556Myocardial InfarctionDeath35.235.25Fyoeradial InfarctionDeath61.561.5HypertensionMyocardial InfarctionS2.568.2Myocardial InfarctionCensoring30.930.96623DiabetesMyocardial Infarction85.886.07623DiabetesMyocardial Infarction85.886.07526Myocardial InfarctionCensoring67.267.27526Myocardial InfarctionDeath30.630.97526Myocardial InfarctionCensoring60.560.5Myocardial InfarctionDeath38.338.338.38Myocardial InfarctionDeath38.338.38Myocardial InfarctionDeath36.558.5952OstoearthritisMyocardial Infarction82.483.99514Myocardial InfarctionDeath36.536.59514Myocardial InfarctionDeath36.636.59525DiabetesMyocardial Infarction86.486.410S25DiabetesMyocardial Infarction86.436.59S25DiabetesMyocardial Infarction50.050.011Myocardial InfarctionCensoring50.050.011Myocardial InfarctionDeath64.262.2 <td>4</td> <td>680</td> <td>Hypertension</td> <td>Myocardial Infarction</td> <td>78.3</td> <td>79.5</td>	4	680	Hypertension	Myocardial Infarction	78.3	79.5
Myocardial InfarctionDeath35.235.25Myocardial InfarctionDeath61.561.56HypertensionMyocardial Infarction52.568.2663DiabetesMyocardial Infarction85.886.0Myocardial InfarctionCensoring67.267.267.2Myocardial InfarctionDeath30.630.630.67526Myocardial InfarctionDeath30.630.67526Myocardial InfarctionDeath38.338.38Myocardial InfarctionDeath38.338.37526Myocardial InfarctionDeath36.558.58Myocardial InfarctionDeath36.536.59582OsteoarthritisMyocardial Infarction82.483.99Myocardial InfarctionCensoring58.558.59Myocardial InfarctionDeath36.536.59514Myocardial InfarctionDeath36.536.510525DiabetesMyocardial Infarction43.643.611525DiabetesMyocardial Infarction57.279.911652AthmaMyocardial Infarction26.950.011652AthmaMyocardial Infarction64.245.211652Nyocardial InfarctionCensoring50.050.011652AthmaMyocardial Infarction64.26			Myocardial Infarction	Censoring	62.0	62.0
5556Myocardial InfarctionDeath61.561.5HypertensionMyocardial Infarction52.568.2Myocardial InfarctionCensoring30.930.96623DiabetesMyocardial Infarction85.886.0Myocardial InfarctionCensoring67.267.267.2HypertensionMyocardial Infarction82.483.086.07526Myocardial InfarctionCensoring60.560.5Myocardial InfarctionCensoring60.560.560.58Myocardial InfarctionCensoring82.483.98582OsteoarthritisMyocardial Infarction82.483.99582OsteoarthritisMyocardial Infarction82.483.99514Myocardial InfarctionCensoring85.558.59514Myocardial InfarctionDeath36.636.610S25DiabetesMyocardial Infarction84.483.911652DiabetesMyocardial Infarction57.279.9Myocardial InfarctionCensoring7.279.911652AthmaMyocardial Infarction50.050.011652AthmaMyocardial Infarction9.846.211652AthmaMyocardial Infarction9.846.212HypertensionCensoring50.050.013Myocardial InfarctionCensoring5			Myocardial Infarction	Death	35.2	35.2
AppertensionMyocardial Infarction52.568.2Myocardial InfarctionCensoring30.930.96DiabetesMyocardial Infarction85.886.0Myocardial InfarctionCensoring67.267.2Myocardial InfarctionCensoring67.268.7Myocardial InfarctionDeath30.630.67526Myocardial InfarctionDeath30.630.67526Myocardial InfarctionDeath38.338.3Myocardial InfarctionDeath37.737.78582OsteoarthritisMyocardial Infarction82.483.9Myocardial InfarctionCensoring58.558.59514Myocardial InfarctionDeath36.536.59514Myocardial InfarctionDeath36.636.510525DiabetesMyocardial Infarction86.486.811652DiabetesMyocardial Infarction50.050.011652AsthmaMyocardial InfarctionCensoring50.050.011652AsthmaMyocardial InfarctionMyocardial Infarction64.246.211652AsthmaMyocardial InfarctionGensoring50.050.011652AsthmaMyocardial InfarctionGensoring52.050.011652AsthmaMyocardial InfarctionGensoring50.262.212Myocardial	5	556	Myocardial Infarction	Death	61.5	61.5
AgeMyocardial InfarctionCensoring30.930.96623DiabetesMyocardial Infarction85.886.0Myocardial InfarctionCensoring67.267.2HypertensionMyocardial Infarction49.268.7Myocardial InfarctionDeath30.630.67526Myocardial InfarctionCensoring60.560.5Myocardial InfarctionDeath38.338.338.38582OsteoarthritisMyocardial Infarction82.483.9Myocardial InfarctionCensoring58.558.558.59514Myocardial InfarctionDeath36.536.59514Myocardial InfarctionDeath51.651.610525DiabetesMyocardial Infarction43.643.611652AsthmaMyocardial InfarctionCensoring50.050.011652AsthmaMyocardial InfarctionDeath46.246.211652Myocardial InfarctionDeath66246.211652AsthmaMyocardial InfarctionPoetnig50.050.011652AsthmaMyocardial InfarctionAsthmaMyocardial Infarction65.265.211652Myocardial InfarctionCensoring76.275.075.0			Hypertension	Myocardial Infarction	52.5	68.2
6623DabetesMyocardial Infarction85.886.0Myocardial InfarctionCensoring67.267.2HypertensionMyocardial Infarction49.268.7Myocardial InfarctionDeath30.630.67526Myocardial InfarctionCensoring60.560.5Myocardial InfarctionDeath38.338.338.38582OsteoarthritisMyocardial Infarction82.483.9Myocardial InfarctionMyocardial Infarction82.483.98582OsteoarthritisMyocardial Infarction82.483.99514Myocardial InfarctionCensoring51.651.69514Myocardial InfarctionDeath36.536.510525DiabetesMyocardial Infarction84.486.811652AsthmaMyocardial Infarction50.050.011652AsthmaMyocardial Infarction94.6246.211Myocardial InfarctionMyocardial Infarction94.898.811Myocardial InfarctionMyocardial Infarction94.898.811Myocardial InfarctionMyocardial Infarction94.898.811Myocardial InfarctionMyocardial Infarction94.898.811Myocardial InfarctionMyocardial Infarction94.898.811Myocardial InfarctionMyocardial Infarction94.6262.212M			Myocardial Infarction	Censoring	30.9	30.9
Myocardial InfarctionCensoring67.267.2HypertensionMyocardial Infarction49.268.7Myocardial InfarctionDeath30.630.67526Myocardial InfarctionCensoring60.560.5Myocardial InfarctionDeath38.338.3Myocardial InfarctionDeath38.338.3Myocardial InfarctionDeath38.338.3Myocardial InfarctionDeath38.338.3Myocardial InfarctionMyocardial Infarction82.483.9Myocardial InfarctionMyocardial Infarction82.483.9Myocardial InfarctionCensoring58.558.5Myocardial InfarctionDeath36.536.59514Myocardial InfarctionDeath36.636.510525DiabetesMyocardial Infarction43.643.611652AsthmaMyocardial InfarctionDeath57.279.911652AsthmaMyocardial Infarction9.846.246.211652AsthmaMyocardial Infarction9.849.846.211652AsthmaMyocardial Infarction9.849.846.211652HypertensionCensoring76.276.211Myocardial InfarctionMyocardial Infarction9.849.811652Myocardial InfarctionMyocardial Infarction9.849.811652 <td< td=""><td>6</td><td>623</td><td>Diabetes</td><td>Myocardial Infarction</td><td>85.8</td><td>86.0</td></td<>	6	623	Diabetes	Myocardial Infarction	85.8	86.0
HypertensionMyocardial Infarction49.268.7Myocardial InfarctionDeath30.630.67526Myocardial InfarctionCensoring60.560.5Myocardial InfarctionDeath38.338.38Myocardial InfarctionDeath37.737.78582OsteoarthritisMyocardial Infarction82.483.99Myocardial InfarctionCensoring58.558.59Myocardial InfarctionDeath36.536.59514Myocardial InfarctionDeath36.536.510525DiabetesMyocardial Infarction43.643.611652AsthmaMyocardial InfarctionCensoring57.279.911652AsthmaMyocardial InfarctionDeath62.260.011Myocardial InfarctionDeath62.260.060.011Myocardial InfarctionCensoring64.262.212Myocardial InfarctionDeath62.260.013Myocardial InfarctionDeath62.260.014Myocardial InfarctionCensoring50.060.015DiabetesMyocardial Infarction64.262.216Myocardial InfarctionDeath62.262.217Myocardial InfarctionDeath64.262.218Myocardial InfarctionDeath64.262.219Myocardial	6		Myocardial Infarction	Censoring	67.2	67.2
Myocardial InfarctionDeath30.630.67526Myocardial InfarctionCensoring60.560.5Myocardial InfarctionDeath38.338.38Myocardial InfarctionHypertension37.737.78582OsteoarthritisMyocardial Infarction82.483.99Myocardial InfarctionCensoring58.558.5914Myocardial InfarctionDeath36.536.59514Myocardial InfarctionDeath51.651.610525DiabetesMyocardial Infarction43.643.611652AsthmaMyocardial Infarction50.050.011652AsthmaMyocardial Infarction0.050.011Myocardial InfarctionCensoring76.276.211Myocardial InfarctionMyocardial Infarction94.898.811Myocardial InfarctionMyocardial Infarction57.276.211Myocardial InfarctionMyocardial Infarction57.276.212Myocardial InfarctionMyocardial Infarction94.898.813Myocardial InfarctionMyocardial Infarction76.276.214Myocardial InfarctionMyocardial Infarction76.276.215Myocardial InfarctionMyocardial Infarction76.276.216Myocardial InfarctionMyocardial Infarction76.276.217Myoc			Hypertension	Myocardial Infarction	49.2	68.7
7526Myocardial InfarctionCensoring60.560.5Myocardial InfarctionDeath38.338.3Myocardial InfarctionHypertension37.737.78582OsteoarthritisMyocardial Infarction82.483.9Myocardial InfarctionCensoring58.558.5HypertensionMyocardial Infarction46.768.09514Myocardial InfarctionDeath36.536.510525DiabetesMyocardial Infarction86.486.811652AsthmaMyocardial Infarction50.050.011652AsthmaMyocardial Infarction94.898.811Myocardial InfarctionCensoring76.276.211Myocardial InfarctionMyocardial Infarction94.898.811Myocardial InfarctionMyocardial Infarction94.898.811Myocardial InfarctionMyocardial Infarction76.276.212Myocardial InfarctionMyocardial Infarction38.475.0			Myocardial Infarction	Death	30.6	30.6
Myocardial InfarctionDeath38.338.3Myocardial InfarctionHypertension37.737.78582OsteoarthritisMyocardial Infarction82.483.9Myocardial InfarctionCensoring58.558.5HypertensionMyocardial Infarction66.768.09514Myocardial InfarctionDeath36.536.59514Myocardial InfarctionDeath51.651.610525DiabetesMyocardial Infarction43.643.611652AsthmaMyocardial Infarction52.450.011652AsthmaMyocardial Infarction0.046.246.211Myocardial InfarctionCensoring76.276.211Myocardial InfarctionDeath52.550.011Myocardial InfarctionDeath65265.211652AsthmaMyocardial Infarction64.212Myocardial InfarctionDeath46.246.213652Myocardial InfarctionDeath46.246.214Myocardial InfarctionDeath46.246.246.215Myocardial InfarctionDeath65.265.265.216Myocardial InfarctionMyocardial Infarction94.898.816Myocardial InfarctionCensoring76.276.217Myocardial InfarctionMyocardial Infarction38.475.0 </td <td>7</td> <td>526</td> <td>Myocardial Infarction</td> <td>Censoring</td> <td>60.5</td> <td>60.5</td>	7	526	Myocardial Infarction	Censoring	60.5	60.5
Mocardial InfarctionHypertension37.737.78582OsteoarthritisMyocardial Infarction82.483.9Myocardial InfarctionCensoring58.558.5HypertensionMyocardial Infarction66.768.09514Myocardial InfarctionDeath36.536.510525DiabetesMyocardial Infarction57.279.911652AsthmaMyocardial InfarctionDeath60.050.011652AsthmaMyocardial InfarctionMyocardial Infarction94.898.811Myocardial InfarctionMyocardial Infarction10.076.276.2			Myocardial Infarction	Death	38.3	38.3
8582OsteoarthritisMyocardial Infarction82.483.9Myocardial InfarctionCensoring58.558.5HypertensionMyocardial Infarction46.768.09514Myocardial InfarctionDeath36.536.59514Myocardial InfarctionDeath51.651.610525DiabetesMyocardial Infarction43.643.611652AsthmaMyocardial Infarction50.050.011652AsthmaMyocardial Infarction98.811652Myocardial InfarctionCensoring76.276.211652Myocardial InfarctionCensoring53.050.011652AsthmaMyocardial Infarction38.475.0			Myocardial Infarction	Hypertension	37.7	37.7
Myocardial InfarctionCensoring58.558.5HypertensionMyocardial Infarction46.768.09514Myocardial InfarctionDeath36.536.59514Myocardial InfarctionDeath51.651.610Ayocardial InfarctionCensoring47.547.510525DiabetesMyocardial Infarction86.486.811652Aytocardial InfarctionCensoring50.050.011652AsthmaMyocardial Infarction94.898.811Kocardial InfarctionCensoring62.276.211Myocardial InfarctionCensoring54.357.211Kocardial InfarctionCensoring76.276.211Kocardial InfarctionCensoring76.276.211Kocardial InfarctionCensoring76.276.211Kocardial InfarctionCensoring76.276.212Myocardial InfarctionCensoring76.276.213Kocardial InfarctionKocardial Infarction75.075.0	8	582	Osteoarthritis	Myocardial Infarction	82.4	83.9
Auge and a set of the set of			Myocardial Infarction	Censoring	58.5	58.5
9514Myocardial Infarction Myocardial InfarctionDeath36.536.59514Myocardial InfarctionDeath51.651.6Myocardial InfarctionCensoring47.547.510525DiabetesMyocardial Infarction86.486.810525DiabetesMyocardial Infarction57.279.9Myocardial InfarctionCensoring50.050.050.011652AsthmaMyocardial Infarction94.898.811Kocardial InfarctionCensoring76.276.211Myocardial InfarctionCensoring76.276.211Myocardial InfarctionMyocardial Infarction58.475.0			Hypertension	Myocardial Infarction	46.7	68.0
9514Myocardial InfarctionDeath51.651.6Myocardial InfarctionCensoring47.547.5Myocardial InfarctionHypertension43.643.610525DiabetesMyocardial Infarction86.486.8HypertensionMyocardial InfarctionMyocardial Infarction57.279.9Myocardial InfarctionCensoring50.050.011652AsthmaMyocardial Infarction94.898.8Myocardial InfarctionCensoring76.276.2HypertensionMyocardial InfarctionS8.475.0			Myocardial Infarction	Death	36.5	36.5
Myocardial InfarctionCensoring47.547.5Myocardial InfarctionHypertension43.643.610525DiabetesMyocardial Infarction86.486.8HypertensionMyocardial Infarction57.279.9Myocardial InfarctionCensoring50.050.011652AsthmaMyocardial Infarction46.246.2Myocardial InfarctionCensoring76.276.2HypertensionMyocardial InfarctionSensoring76.276.2	9	514	Myocardial Infarction	Death	51.6	51.6
Nyocardial InfarctionHypertension43.643.610525DiabetesMyocardial Infarction86.486.8HypertensionMyocardial Infarction57.279.9Myocardial InfarctionCensoring50.050.011652AsthmaMyocardial Infarction46.2Myocardial InfarctionCensoring94.898.8Myocardial InfarctionCensoring76.276.2HypertensionMyocardial Infarction38.475.0			Myocardial Infarction	Censoring	47.5	47.5
10525DiabetesMyocardial Infarction86.486.8HypertensionMyocardial Infarction57.279.9Myocardial InfarctionCensoring50.050.0Myocardial InfarctionDeath46.246.211652AsthmaMyocardial Infarction94.898.8Myocardial InfarctionCensoring76.276.2HypertensionMyocardial Infarction38.475.0			Myocardial Infarction	Hypertension	43.6	43.6
HypertensionMyocardial Infarction57.279.9Myocardial InfarctionCensoring50.050.011652AsthmaDeath46.246.2Myocardial InfarctionKensoring94.898.8Myocardial InfarctionCensoring76.276.2HypertensionMyocardial Infarction38.475.0	10	525	Diabetes	Myocardial Infarction	86.4	86.8
Myocardial InfarctionCensoring50.050.0Myocardial InfarctionDeath46.246.211652AsthmaMyocardial Infarction94.898.8Myocardial InfarctionCensoring76.276.2HypertensionMyocardial Infarction38.475.0			Hypertension	Myocardial Infarction	57.2	79.9
Nyocardial InfarctionDeath46.246.211652AsthmaMyocardial Infarction94.898.8Myocardial InfarctionCensoring76.276.2HypertensionMyocardial Infarction38.475.0			Myocardial Infarction	Censoring	50.0	50.0
11652AsthmaMyocardial Infarction94.898.8Myocardial InfarctionCensoring76.276.2HypertensionMyocardial Infarction38.475.0			Myocardial Infarction	Death	46.2	46.2
Myocardial InfarctionCensoring76.276.2HypertensionMyocardial Infarction38.475.0	11	652	Asthma	Myocardial Infarction	94.8	98.8
Hypertension Myocardial Infarction 38.4 75.0			Myocardial Infarction	Censoring	76.2	76.2
			Hypertension	Myocardial Infarction	38.4	75.0
Depression Myocardial Infarction 34.3 72.8			Depression	Myocardial Infarction	34.3	72.8

## Table 4 Main sequences of length 2 accross clusters

<sup>a</sup> Number of sequences of size 2 observed in the corresponding cluster

<sup>b</sup> Proportion of patients presenting the sequence

<sup>c</sup> Proportion of patients diagnosed with the second long-term condition given the first condition has been diagnosed

of mental health conditions. Non-conventional risk factors associated with these clusters were peripheral vascular disease and COPD (cluster 7) and atrial fibrillation (cluster 9).

Association between these non-conventional risk factors and myocardial infarction have been previously reported with various levels of evidence [31–35]. Conventional assumptions regarding these possible associations include i) the shared risk factors hypotheses [31, 36], ii) the complications induced by long-term exposures to pharmacotherapies used in the treatment of these LTCs, such as non-steroidal anti-inflammatories

used in the treatment of ostheoarthritis [37], beta-2 agonists and systemic corticosteroids used in the treatment of asthma [38–40] and antipsychotic drugs used in the treatment of mental health disorders [41], and iii) induced restrictions on physical activity [36], which in return are associated with increased levels of CVRFs such as hypercholesterolaemia, hypertension and diabetes.

Examination of the typology created using *multiple* state analysis shows that resulting clusters of patients were discriminated not only in terms of the main myocardial infarction driving LTCs but also with respect to major socio-demographics variables including age, gender and ethnicity, as well as other CVRFs such as hypercholesterolaemia or ever smoking status. Therefore, assuming that different early life lifestyle and exposures, possibly mediated by genotype and environment factors, result in distinct patterns of multimorbidity, tractable through the continuous record of health conditions, our results support both the environmental and the genotype / environment hypothesis of life course epidemiology. From a public health perspective however, underlying models are not of prime importance as public health policies would rather focus on the modifiable risk factors. To this regard, although multiple state clustering analysis presents itself as a relevant methodological approach to electronic health records-allowing patients' stratification in terms of both the main driving LTCs as well as associated socio-demographic variables-further research is needed. This includes a deeper assessment of typologies resulting from this methodology, particularly in therms of their clinical and public health utility, to better evaluate its potential for for informing patientoriented health policies.

#### Abbreviations

LTC	Long-term condition
MSCA	Multiple State Analysis
CVRF	Cardiovascular risk factor
IMD	Index of multiple deprivation

#### Acknowledgements

We dedicate this work to the memory of Professor Mark Ashworth, whose unwavering support and encouragement profoundly shaped our efforts.

#### Authors' contributions

Conceptualization: Abdel Douiri and Marc Delord; Formal analysis and investigation: Marc Delord; Writing - original draft preparation: Marc Delord; Writing - review and editing: Abdel Douiri; Funding acquisition: Abdel Douiri; Supervision: Abdel Douiri. Both authors approved the final manuscript.

#### Funding

This project is funded by King's Health Partners / Guy's and St Thomas Charity 'MLTC Challenge Fund' (grant number ElC180702) and support from the National Institute for Health and Care Research (NIHR) under its Programme Grants for Applied Research (NIHR202339).

#### Data availability

Restrictions apply to the availability of the data supporting the findings of this study. The data were used under license, following project approval by the Lambeth Clinical Commissioning Group, and therefore are not publicly available. The R library MSCA, which implements user-friendly routines for performing MSCA, will soon be available in the CRAN repository (https:// cran.r-project.org/).

## Declarations

#### Ethics approval and consent to participate

This study was conducted in accordance with the Declaration of Helsinki. Ethics approval was granted by the Health Research Authority (HRA) ethics committees (IRAS: 282174) and the Confidentiality Advisory Group (CAG: 22/ CAG/0022), as well as the local Research Ethics Committees (RECs) at Guy's and St Thomas' NHS Foundation Trust (London, UK), King's College Hospital NHS Foundation Trust (London, UK), Queen Square (London, UK), and Westminster Hospital (London, UK) (REC: 22/SC/0043).

#### **Consent for publication**

Not applicable.

#### **Competing interests**

The authors declare no competing interests.

Received: 30 June 2023 Accepted: 17 January 2025 Published online: 04 February 2025

#### References

- 1. Elder GH Jr. The life course as developmental theory. Child Dev. 1998;69(1):1–12.
- 2. Hill AB, Doll R. Smoking habits of doctors. Br Med J. 1957;2(5054):1173.
- Rutter M, Sroufe LA. Developmental psychopathology: Concepts and challenges. Dev Psychopathol. 2000;12(3):265–96.
- 4. Brown GW, Harris T. Social origins of depression: A study of psychiatric disorder in women. J Psychiatr Res. 1978;12(3):189.
- Younossi ZM, Koenig AB, Abdelatif D, Fazel Y, Henry L, Wymer M. Global epidemiology of nonalcoholic fatty liver disease-meta-analytic assessment of prevalence, incidence, and outcomes. Hepatology. 2016;64(1):73–84.
- Herrett E, Gallagher AM, Bhaskaran K, Forbes H, Mathur R, van Staa T, Smeeth L. Data resource profile: Clinical practice research datalink (cprd). Int J of Epidemiol. 2015;44(3):827–36.
- Carter BD, Abnet CC, Feskanich D, Freedman ND, Hartge P, Lewis CE, Ockene JK, Prentice RL, Speizer FE, Thun MJ, et al. Smoking and mortalitybeyond established causes. N Engl J Med. 2015;372(7):631–40.
- Roth GA, Huffman MD, Moran AE, Feigin V, Mensah GA, Naghavi M, Murray CJL. Global and regional patterns in cardiovascular mortality from 1990 to 2013. Circulation. 2015;132(17):1667–78.
- Jaffe AS. Third universal definition of myocardial infarction. Clin Biochem. 2013;46(1–2):1–4.
- 10. Libby P. Mechanisms of acute coronary syndromes and their implications for therapy. N Engl J Med. 2013;368:2004–13.
- Yusuf PS, Hawken S, Ôunpuu S, Dans T, Avezum A, Lanas F, McQueen M, Budaj A, Pais P, Varigos J, Lisheng L. Effect of potentially modifiable risk factors associated with myocardial infarction in 52 countries (the interheart study): case-control study. Lancet. 2004;364:937–52.
- Anand SS, Islam S, Rosengren A, Franzosi MG, Steyn K, Yusufali AH, Keltai M, Diaz R, Rangarajan S, Yusuf S, I. N. T. E. R. H. E. A. R. T. Investigators. Risk factors for myocardial infarction in women and men: insights from the interheart study. Eur Heart J. 2008;29:932–40.
- Nielsen M, Andersson C, Gerds TA, Andersen PK, Jensen TB, Køber L, Gislason G, Torp-Pedersen C. Familial clustering of myocardial infarction in first-degree relatives: a nationwide study. Eur Heart J. 2013;34:1198–203.
- 14. Kaufman L, Rousseeuw PJ. Finding Groups in Data: An Introduction to Cluster Analysis. 1st ed. Hoboken: Wiley; 2009.

- 15. Jaccard P. Étude comparative de la distribution florale dans une portion des alpes et des jura. Bull Soc Vaudoise Sci Nat. 1901;37:547–79.
- Ward JH Jr. Hierarchical grouping to optimize an objective function. J Am Stat Assoc. 1963;58(301):236–44.
- Murtagh F, Legendre P. Ward's hierarchical agglomerative clustering method: which algorithms implement ward's criterion? J Classif. 2014;31(3):274–95.
- Lance GN, Williams WT. A general theory of classificatory sorting strategies: 1. hierarchical systems. Comput J. 1967;9(4):373–80.
- Müllner D. fastcluster: Fast hierarchical, agglomerative clustering routines for r and python. J Stat Softw. 2013;53(9):1–18.
- Müllner D. Modern hierarchical, agglomerative clustering algorithms. arXiv preprint arXiv:1109.2378. 2011.
- 21. Milligan GW, Cooper MC. An examination of procedures for determining the number of clusters in a data set. Psychometrika. 1985;50:159–79.
- Sugiyama K, Tagawa S, Toda M. Methods for visual understanding of hierarchical system structures. IEEE Trans Syst Man Cybern. 1981;11(2):109–25.
- 23. R Core Team. R: A Language and Environment for Statistical Computing. Vienna: R Foundation for Statistical Computing; 2023.
- Marc Delord. MSCA: Clustering of multiple censored time-to-event endpoints, 2025. R package version 1.0 (beta version). Available at https:// cran.r-project.org/package=MSCA.
- Eisen MB, Spellman PT, Brown PO, Botstein D. Cluster analysis and display of genome-wide expression patterns. Proc Natl Acad Sci. 1998;95(25):14863–8.
- Pollock G. Holistic trajectories: a study of combined employment, housing and family careers by using multiple-sequence analysis. J R Stat Soc Ser A (Stat Soc). 2007;170(1):167–83.
- Gauthier J-A, Widmer ED, Bucher P, Notredame C. Multichannel sequence analysis applied to social science data. Sociol Methodol. 2010;40(1):1–38.
- Barnett K, Mercer SW, Norbury M, Watt G, Wyke S, Guthrie B. Epidemiology of multimorbidity and implications for health care, research, and medical education: a cross-sectional study. Lancet. 2012;380:37–43.
- Ng SK, Tawiah R, Sawyer M, Scuffham P. Patterns of multimorbid health conditions: a systematic review of analytical methods and comparison analysis. Int J Epidemiol. 2018;47(5):1687–704.
- Delord M, Sun X, Learoyd A, Curcin V, Wolfe C, Ashworth M, Douiri A. Patient-oriented unsupervised learning to uncover the patterns of multimorbidity associated with stroke using primary care electronic health records. BMC Prim Care. 2024;25(1):419.
- Hall AJ, Stubbs B, Mamas MA, Myint PK, Smith TO. Association between osteoarthritis and cardiovascular disease: systematic review and metaanalysis. Eur J Prev Cardiol. 2016;23(9):938–46.
- Iribarren C, Tolstykh IV, Miller MK, Sobel E, Eisner MD. Adult asthma and risk of coronary heart disease, cerebrovascular disease, and heart failure: a prospective study of 2 matched cohorts. Am J Epidemiol. 2012;176(11):1014–24.
- Bang DW, Wi CI, Kim EN, Hagan J, Roger V, Manemann S, Lahr B, Ryu E, Juhn YJ. Asthma status and risk of incident myocardial infarction: a population-based case-control study. J Allergy Clin Immunol Pract. 2016;4(5):917–23.
- Nicholson A, Kuper H, Hemingway H. Depression as an aetiologic and prognostic factor in coronary heart disease: a meta-analysis of 6362 events among 146 538 participants in 54 observational studies. Eur Heart J. 2006;27(23):2763–74.
- Roest AM, Martens EJ, de Jonge P, Denollet J. Anxiety and risk of incident coronary heart disease: a meta-analysis. J Am Coll Cardiol. 2010;56(1):38–46.
- Stubbs B, Hurley M, Smith T. What are the factors that influence physical activity participation in adults with knee and hip osteoarthritis? a systematic review of physical activity correlates. Clin Rehabil. 2015;29(1):80–94.
- 37. Schmidt M. Cardiovascular risks associated with non-aspirin non-steroidal anti-inflammatory drug use. Dan Med J. 2015;62:pii–B4987.
- Appleton SL, Ruffin RE, Wilson DH, Taylor AW, Adams RJ, North West Adelaide Cohort Health Study Team, et al. Cardiovascular disease risk associated with asthma and respiratory morbidity might be mediated by short-acting β2-agonists. J Allergy Clin Immunol. 2009;123(1):124–30.
- Zeiger R, Sullivan P, Chung Y, Kreindler JL, Zimmerman NM, Tkacz J. Systemic corticosteroid-related complications and costs in adults with persistent asthma. J Allergy Clin Immunol Pract. 2020;8(10):3455–65.

- 40. Cazzola M, Rogliani P, Calzetta L, Matera MG. Bronchodilators in subjects with asthma-related comorbidities. Respir Med. 2019;151:43–8.
- 41. Nielsen RE, Banner J, Jensen SE. Cardiovascular disease in patients with severe mental illness: a review. Nat Rev Cardiol. 2021;18(2):136–145.

## **Publisher's Note**

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.