

OPEN

# MorbiNet: multimorbidity networks in adult general population. Analysis of type 2 diabetes mellitus comorbidity

Alba Aguado<sup>1\*</sup>, Ferran Moratalla-Navarro<sup>2,3,4,5</sup>, Flora López-Simarro<sup>6</sup> & Victor Moreno<sup>2,3,4,5\*</sup>

Multimorbidity has great impact on health care. We constructed multimorbidity networks in the general population, extracted subnets focused on common chronic conditions and analysed type 2 diabetes mellitus (T2DM) comorbidity network. We used electronic records from 3,135,948 adult people in Catalonia, Spain (539,909 with T2DM), with at least 2 coexistent chronic conditions within the study period (2006–2017). We constructed networks from odds-ratio estimates adjusted by age and sex and considered connections with OR > 1.2 and p-value < 1e-5. Directed networks and trajectories were derived from temporal associations. Interactive networks are freely available in a website with the option to customize characteristics and subnets. The more connected conditions in T2DM undirected network were: complicated hypertension and atherosclerosis/peripheral vascular disease (degree: 32), cholecystitis/cholelithiasis, retinopathy and peripheral neuritis/neuropathy (degree: 31). T2DM has moderate number of connections and centrality but is associated with conditions with high scores in the multimorbidity network (neuropathy, anaemia and digestive diseases), and severe conditions with poor prognosis. The strongest associations from T2DM directed networks were to retinopathy (OR: 23.8), glomerulonephritis/nephrosis (OR: 3.4), peripheral neuritis/neuropathy (OR: 2.7) and pancreas cancer (OR: 2.4). Temporal associations showed the relevance of retinopathy in the progression to complicated hypertension, cerebrovascular disease, ischemic heart disease and organ failure.

Multimorbidity is the simultaneous presence of two or more chronic medical conditions<sup>1</sup>. It is a common problem, more than one-third of the patients visited by primary care physicians have four or more chronic health problems and a small proportion has more than ten during their life<sup>2</sup>. The prevalence is higher in older people and low socioeconomic situations<sup>3</sup>. It is associated with a poorer quality of life, more disability<sup>4</sup> and patient safety incidents<sup>5</sup> and a greater, almost exponential, increase in health care costs<sup>6</sup>.

Multimorbidity has raised increasing interest in the last years. It is a complex phenomenon and can be studied with network analysis. Hidalgo *et al.* studied phenotypic networks and found that patients with diseases highly connected tend to die sooner<sup>7</sup>. This tool has been used to identify comorbidity associated with hypertension<sup>8</sup>, chronic pulmonary obstructive disease<sup>9</sup> or mental disease<sup>10</sup> and to compare multimorbidity by gender<sup>11</sup>. Most of the studies used hospital databases to construct the networks. It could be useful to approach the study of multimorbidity for an individual patient focusing on an initial condition or the one with the highest impact. Thus, comorbidity networks could help to analyse multimorbidity.

There are 425 million people with diabetes in the world and the prevalence is increasing (IDF Atlas) because of the aging population and lifestyle habits<sup>12</sup>. In Spain, the adjusted prevalence for age and sex of diabetes mellitus was estimated to be 13.8% (6.0% being undiagnosed)<sup>12</sup>. Comorbidities are common in patients with T2DM<sup>13</sup> and the proportion of diabetic patients with multimorbidity increases after diagnosis (from 32% to 80% after 16

<sup>1</sup>CAP Sagrada Familia. Consorci Sanitari Integral, Barcelona, Spain. <sup>2</sup>Oncology Data Analytics Program, Catalan Institute of Oncology (ICO), L'Hospitalet de Llobregat, Barcelona, Spain. <sup>3</sup>ONCOBELL Program, Bellvitge Biomedical Research Institute (IDIBELL), L'Hospitalet de Llobregat, Barcelona, Spain. <sup>4</sup>CIBER of Epidemiology and Public Health (CIBERESP), Madrid, Spain. <sup>5</sup>Department of Clinical Sciences, Faculty of Medicine and Health Sciences, University of Barcelona, Barcelona, Spain. <sup>6</sup>ABS Urban Martorell, Catalan Institute of Health, Martorell, Barcelona, Spain. \*email: [alba.aguado@sanitatintegral.org](mailto:alba.aguado@sanitatintegral.org); [v.moreno@iconcologia.net](mailto:v.moreno@iconcologia.net)

years)<sup>14</sup>. Multimorbidity in diabetics is associated with a reduced quality of life<sup>15</sup>, increased cost<sup>16</sup> and mortality<sup>17</sup>. Diseases associated with diabetes have widely been studied, but not using the approach of network analysis.

We aimed to construct the multimorbidity network in the general population, extract subnets focused on the most common chronic health conditions, construct an interactive website openly available to visualize the networks and analyse T2DM comorbidity network.

## Methods

**Study design, population and variables.** We performed a retrospective longitudinal population-based study in adults in Catalonia, Spain. We used electronic health records from the Information System for Research Development in Primary Care (SIDIAP), which includes clinical data from approximately 5.5 million people, 74% of the population in Catalonia. All of them are users of the public health care system.

We included subjects aged 18 years or more with at least two coexistent chronic health conditions active any time within the period 2006–2017. Diseases were coded with the International Statistical Classification of Diseases and Related Health Problems 10th revision (ICD-10) system<sup>18</sup> in the electronic records. ICD-10 code diagnoses were mapped to the International Classification of Primary Care, 2<sup>nd</sup> edition (ICPC-2) system<sup>19</sup>. We followed an adapted version of the list of chronic conditions based on ICPC-2 codes described by O'Halloran<sup>20</sup>. We obtained the variables: date of birth, sex, diagnosis (date of diagnosis and resolution, ICPC-2 code), tobacco and alcohol consumption, socioeconomic group, weight, height and rurality). Patient's age was defined as the mean value between the age when the first and last diagnoses were made within the study period or the age in the middle of the period for the cases without new diagnoses made in the period.

For the T2DM comorbidity analysis, we included the patients with multimorbidity with the ICPC-2 code T90 (diabetes non-insulin dependent), originally coded in ICD-10 as E11 (type 2 diabetes mellitus).

**Network construction and analysis.** We used logistic regression models, adjusted by age and sex, to construct the multimorbidity networks. A relational database was developed on a server to store the variables at the individual level, with a table for patient characteristics and another one for multiple diagnoses records. We used free R software and R Igraph library to prepare database, construct the networks and obtain network parameters and node attributes. The conditions with less than 1000 patients were not included in the networks. For the patients who died or transferred to a different region, we included the data until the date of death or transfer.

The basic elements of the networks are nodes (chronic health conditions) and edges (coexistence of disease) that connect nodes within the network. Graphs are visual representations of networks, with nodes and edges. Connections are defined according to the criteria of coexistence to a greater degree than expected by the prevalence of diseases. For these criteria, thresholds were used in the association measures (odds ratios (OR)  $\geq 1.2$ ) and p-value  $< 1e-5$  (Bonferroni correction to account for multiple tests). These filters we applied because the large sample size analysed would generate significant associations even for very small magnitudes. Edge thickness in the network plots varied proportional to OR values. We constructed non-directional networks, and also directed networks and trajectories for the specific analysis of temporal associations.

**Directed networks and trajectories.** We constructed directed networks to assess temporal disease associations. We identified sequential associations among pairs of diagnoses to study temporal patterns. We only considered a temporal association for probabilities below 40% or above 60% that a disease was diagnosed previously or afterward another one. Conditions diagnosed between 41 and 59% of the times before or after another one were considered to have no temporal association. Thus, a diagnosis  $D_a$  was considered to precede diagnosis  $D_b$  when in at least 60% of patients it took place before (sequence  $D_a - D_b$ ). When  $D_a$  was made before  $D_b$  in less than 40% of the patients we assumed the sequence was  $D_b - D_a$ . This approach might be useful to analyse disease progression and study temporal directionality in multimorbidity.

With the directed networks for a specific disease, we constructed a trajectory. We considered only directed associations with odds ratios  $\geq 1.5$ . The trajectory was made with the conditions connected to the disease of interest (first steps before it) and other chronic diseases connected to (up to three steps from the disease of interest).

**Node attributes and network parameters.** We obtained relevant node attributes such as the degree, clustering coefficient and PageRank index. The *degree* ( $K_i$ ) measures the connectivity of the node<sup>21</sup>, as the number of links connected to node  $i$ . The *clustering coefficient* measures the likelihood that two nodes connected to node  $i$  are connected themselves. It is a density measure of local connections<sup>22</sup> and shows the tendency of the network to aggregate in subgroups. It's a number between 0 and 1, and is calculated as<sup>21</sup>:

$$\text{Clustering coefficient}_i = \frac{2e_i}{k_i(k_i - 1)}$$

Where  $e_i$  is the number of existing links among the  $k_i$  nodes that connect to node  $i$

The *pageRank* measures node influence, based on the number of links it has to other nodes in the network and the links their connections have, taking into account links direction and weight.

In a directed network PageRank index for disease  $g$ ,  $PR(g)$ , is calculated as<sup>23</sup>:

$$PR(g) = (1 - d)N^{-1} + \sum_{u \in U(g)} PR(u)/N_{ds}(u).$$

where  $U(g)$  is the set of inbound conditions of disease  $g$ ,  $N_{ds}(u)$  is the number of outbound diseases of disease  $u$  and  $N$  is the total number of chronic conditions in the underlying network and  $d$  is a damping factor that has been set to 0.85.

In an undirected weighted network, the edges are considered as bi-directional and the classic centralities definition is applied.

Regarding the whole network, we calculated the number of nodes and edges, the clustering coefficient, diameter, density and centralization:

The *Number of nodes* is the number of network units and the *Number of edges* the number of links connecting the nodes.

The network *clustering coefficient* is the average of the clustering coefficients for all nodes in the network. It measures the degree to which nodes in a graph tend to cluster together.

*Network diameter* is the maximum distance between any two nodes in a network<sup>21</sup>:

$$\text{Network Diameter} = \max\{d_{ij}|i, j \in N\}$$

where  $N$  is the set of nodes in the network.

The *network density* is defined as the number of connections between nodes divided by the number of possible ties or connections<sup>22</sup>. It is a measure of the relative number of connections. It is a value between 0 and 1. A network with only isolated nodes has a density of 0. It is calculated as:

$$\text{Density} = \frac{\sum_i \sum_j m_{ij}}{n(n-1)}$$

where  $M = [m_{ij}]$  is the adjacency  $n \times n$  matrix

*Centralization* is the extent to which network links are focused on one or a few nodes<sup>24</sup>. It is the difference between the number of links for each node divided by the maximum possible sum of differences. It indicates the degree to which a network approaches the configuration of a star network and is related to the degree of asymmetry. Centralization is expressed as a percentage, and it has values from 0 (every node is connected to every other node) to 100 (all nodes are connected to only one node). A centralized network will have many of its links around one or a few nodes.

**Subnets focused on chronic conditions.** The most common chronic conditions were zoomed, and subnets extracted from the general multimorbidity network. We analysed here the T2DM subnet with ICPC-2 codes and  $OR \geq 1.2$ .

**Quality of data and sensitivity analyses.** Since this study was based on routinely collected data, and that can vary from doctor to doctor, we selected the group of patients who had recorded in their electronic health record the variables tobacco and alcohol consumption, weight and height, as a measure of better-quality records. The subgroup of patients with all these variables correctly recorded was 2,214,388 (70% from total) and was compared to the patients with any of them missing. We studied the correlations for the prevalence of diseases in both groups and only for the more infrequent chronic conditions the differences were slightly lower (for the different diseases  $r^2$  ranged from 0.94 to 0.98). The prevalence of most chronic conditions was similar in both groups. Because of this, we used the data of the entire population to construct the networks. Patients with a higher burden of disease would be expected to have more clinical information recorded. For health problems where alcohol, tobacco or weight are risk factors this information might be more carefully obtained. Older patients have more morbidity, are visited more often and have more chances to have this data in their health records. If we would have included only the patients with better quality records, we might be selecting the population with more diseases and introducing bias to the study.

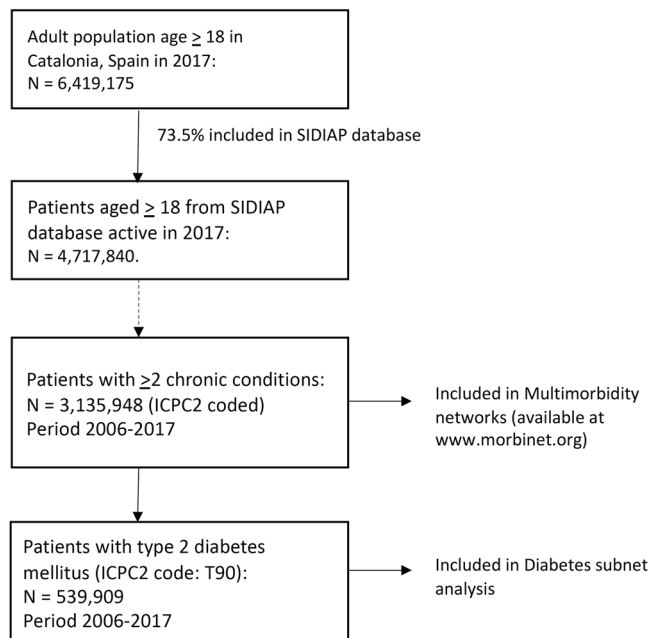
As sensitivity analyses, we analysed how T2DM ICPC-2 networks changed when we modified several conditions (threshold OR value, including only the patients with more complete data versus all patients, changing the criteria to define directionality with probabilities  $<20\%$  or  $>80\%$  of previous/subsequent diagnosis among pairs of conditions versus  $<40\%$  or  $>60\%$ ).

**Website.** The networks and results of the study can be freely accessed on an open website<sup>25</sup>. We have developed an application where users can visualize interactively different networks according to the selection of characteristics of interest. We used Shiny free software (shiny.rstudio.com), R visNetwork library and R Igraph library to visualize networks and obtained descriptive and topological parameters.

The networks can be viewed with different formats to analyse the relationships among diseases. It is possible to select the disease classification system: ICD-10 with 3 digits, ICPC-2 and a simplified ICPC-2 version listed in Supplementary Table S1. It is also possible to select the minimum odds ratio considered to accept a risk association is present (ranging from 1.2 to 2) and to plot the global multimorbidity network or zoom the subnet of specific chronic conditions. It is also possible to represent protective interactions, with a maximum OR to accept them ranging from 0.8 to 0.5.

The node size shown in plots depends on the degree of the condition in the network and the node colour can be selected according to different criteria: clustering coefficient, pageRank value, system location (such as cardiovascular, respiratory...). For a chronic condition subnet, users can select the colour and the size of the nodes according to the selected attribute value in the general network or within the subnet.

The network parameters can be visualized and downloaded. These include the number of nodes and edges, diameter, shortest pathway, density, average neighbours, clustering coefficient and centralization, both for the general network or subnetwork. The degree for each node and OR for each pair of nodes can also be visualized and downloaded.



**Figure 1.** Study population. We included the information of 3,135,948 adult patients with multimorbidity (539,909 of them with T2DM) from SIDIAP database, which contains clinical data from 73.5% of population in Catalonia, Spain.

Directed networks (or subnetworks) can also be plotted for the selected disease classification system and OR value and it is possible to visualize the network descriptive parameters and node attributes.

The website also includes a tab to obtain descriptive statistics of the study population with some filtering options: gender, age group, diagnose period, tobacco use, region, rurality, and socioeconomic group.

**Ethical issues.** We obtained the approval from Consorci Sanitari Integral Ethics Committee (16/457) and IDIAP Jordi Gol Ethics Committee (P17/207) to conduct the study. IDIAP Jordi Gol provided the data and Consorci Sanitari Integral is the institution to which the first author is affiliated and that received funding to perform the research.

Because of the observational, retrospective and population-based design, it was not possible to obtain the informed consent from patients to use their clinical information. The data used for the analysis were completely anonymized with no variables that might allow identification of individual patients. We followed the Spanish law on the protection of personal data (LOPD 15/1999 of December 14).

## Results

We included the information of 3,135,948 patients with multimorbidity; 539,909 of them with T2DM (Fig. 1). We used the data from 2006 to 2017 to study disease associations and construct the networks, but we used 2017 information to calculate prevalences. In 2017 the prevalence of multimorbidity in patients aged 18 and over was 52.8% for chronic conditions coded with the ICPC-2 system. In Supplementary Table S2 we present it for different age and sex groups. It was higher in women for all age groups and it increased with age: 94.2% of patients aged 70 or more have multimorbidity.

The general multimorbidity undirected network using ICD-10 code system with 3 digits and minimum OR 1.2 had 323 nodes and 7159 edges, using ICPC-2 system it had 148 nodes and 2766 edges (Supplementary Table S3) and with the simplified ICPC-2 version it had 111 nodes and 1607 edges. Multimorbidity networks and subnetworks for specific chronic conditions can be visualized at the open website<sup>25</sup>.

The total prevalence of T2DM (ICPC-2 code: T90) in 2017 was 7.9% (8.8% in men and 6.9% in women). In the study period T2DM was present in 20.7% of men and 14.4% of women with multimorbidity (Supplementary Table S2). This percentage varied according to the age group considered and was highest in older patients and lower socioeconomic groups, though it was similar in rural and urban areas. In men with T2DM, ex-smoker was the more frequent category in tobacco use while most diabetic women were non-smokers. Tobacco use was still present in 24.15% of men and 7.4% of women with T2DM (Supplementary Table S4).

Figure 2 represents T2DM (T90 ICPC-2 code) comorbidity undirected network with chronic conditions associated with an OR  $\geq 1.2$ . It had 38 nodes and 448 edges (see network parameters in Supplementary Table S3). The degrees of the nodes in T2DM subnets are listed in Table 1. The conditions with more connections in non-directed networks were hypertension complicated and atherosclerosis/peripheral vascular disease (both with degree 32), and cholecystitis/cholelithiasis, retinopathy and peripheral neuritis (all with degree 31). The OR values for all pairs of connected nodes are shown in Supplementary Table S5.



**Figure 2.** Type 2 diabetes mellitus comorbidity undirected network ( $OR > 1.2$ ). It represents chronic conditions associated with an  $OR > 1.2$ . The node size shown is proportional to the degree and the colour to the pageRank index (both in the global multimorbidity network). The edges thickness is related to the  $OR$  value for the conditions connected. Thus, a big node with a dark colour has many connections and is very central in the general multimorbidity network. Abbreviations: Dis: disease; Ca: cancer; Un: unspecified. T1D: type 1 diabetes mellitus, Isch: ischemic; Ac Myoc Infarct: acute myocardial infarction; Chron: chronic; Pulm: pulmonary; HT comp: hypertension complicated; Trans: transient; ATC: atherosclerosis; fail: failure; Atr.: atrial; deg: degeneration. Figure created with the MorbiNet application v2.2<sup>25</sup>, <https://morbinet.org/shiny>.

The degree, PageRank and position the nodes of T2DM subnet have in the general multimorbidity network (calculated with all diseases), are shown in Supplementary Table S6. Some nodes of the T2DM subnet, such as peripheral neuritis/neuropathy, anaemia unspecified and diseases of digestive system, have the highest number of connections and are very central in the global multimorbidity network.

Figure 3a represents the directed T2DM network, considering the temporal associations, filtered to show only chronic conditions with an  $OR \geq 1.5$ . It had 23 nodes and 221 edges (Supplementary Table S3). In the cases where the temporal direction is not clear, the edges are represented with a double arrow (in both directions). The edges with undefined direction are counted twice and for this reason, the number of edges is higher in directed networks as compared to undirected for the same minimum  $OR$  value. Table 2 includes the  $OR$  values for the nodes with direct links with T2DM in directed networks. The strongest associations from T2DM were to: retinopathy ( $OR$ : 23.83), glomerulonephritis/nephrosis ( $OR$ : 3.44), peripheral neuritis/neuropathy ( $OR$ : 2.70) and pancreas cancer ( $OR$ : 2.42). T2DM received the strongest connections from obesity/overweight ( $OR$ : 2.58) and ischemic heart disease with angina (1.89). The  $OR$  values for all pairs of connected nodes in each direction are shown in Supplementary Table S7. The nodes in Fig. 3b are the conditions with an  $OR \geq 2$  in a directed network. This network restricted to stronger associations had 11 nodes and 46 edges (Supplementary Table S3). Table 1 also

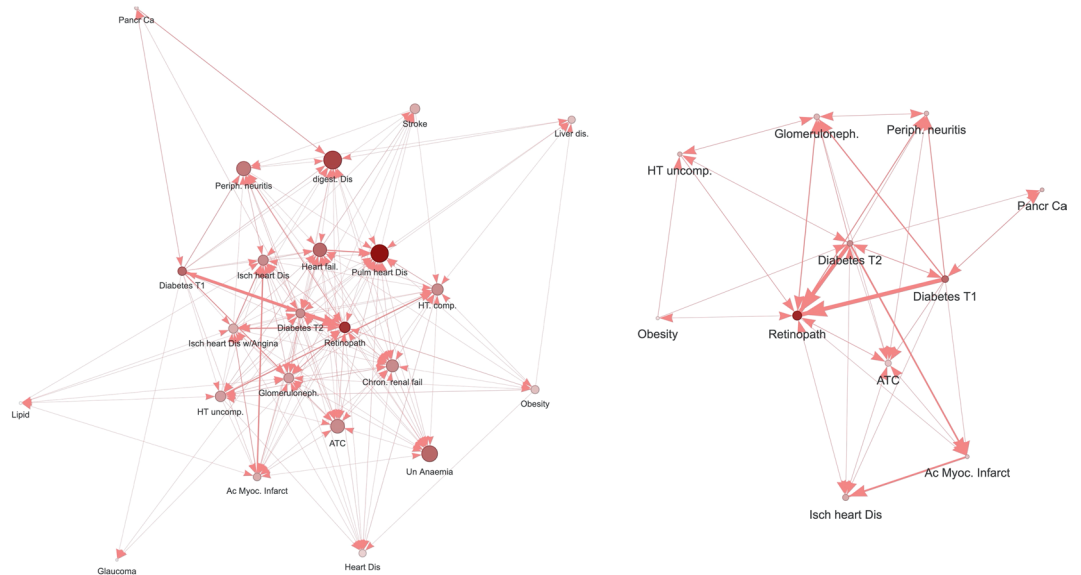
		Non-Directed OR $\geq 1.2$ , Fig. 2	Directed OR $\geq 1.5$ , Fig. 3a		Directed OR $\geq 2$ , Fig. 3b	
		degree	in-degree	out-degree	in-degree	out-degree
T90	Type 2 diabetes mellitus	37	12	20	4	9
K87	Hypertension complicated	32	12	8		
K92	Atherosclerosis/peripheral vascular dis	32	14	13	7	4
D98	Cholecystitis/cholelithiasis	31				
F83	Retinopathy	31	11	16	8	6
N94	Peripheral neuritis/neuropathy	31	11	10	5	3
B82	Anaemia other, unspecified	30	12	10		
U88	Glomerulonephritis/nephrosis	30	16	13	6	4
F92	Cataract	29				
K74	Ischaemic heart dis. with angina	29	8	13		
K76	Ischaemic heart dis. without angina	29	13	9	4	3
U99.01	Chronic renal failure	29	15	10		
D99	Dis. digestive system, other	28	14	7		
K86	Hypertension uncomplicated	28	7	14	3	3
K77	Heart failure	27	15	10		
K84	Heart dis. other	27	7	6		
K91	Cerebrovascular dis.	27				
K89	Transient cerebral ischaemia	26				
K82	Pulmonary heart dis.	25	17	6		
B81	Anaemia vitamin B12/folate deficiency	24				
D97	Liver dis. NOS	24	5	5		
K75	Acute myocardial infarction	24	11	10	4	2
K78	Atrial Fibrillation/flutter	24				
T99	Endocrine/metabolic/nutrition dis. other	24				
T83/82	Obesity/overweight	23	3	10	1	3
T92	Gout	23				
K90	Stroke	22	7	8		
T89	Type 1 diabetes mellitus	22	2	15	2	8
F84	Macular degeneration	21				
P71	Psychosis organic, other	21				
T93	Lipid disorder	21	4	6		
N93	Carpal tunnel syndrome	15				
F93	Glaucoma	14	3	0		
D77	Digestive ca. other/NOS	10				
X77	Female genital ca., other	10				
D76	Pancreas ca.	7	2	2	2	1
B78	Hereditary haemolytic anaemia	5				
P72	Schizophrenia	4				

**Table 1.** Degree of nodes in type 2 diabetes mellitus subnets. Abbreviations: Dis: disease, Ca: cancer, NOS: not otherwise specified.

included the degree of nodes in the directed T2DM subnets (in-degree or number of connections received and out-degree or connections from the considered node), both with  $OR \geq 1.5$  and  $OR > 2$ .

In Fig. 4 we present the trajectories for T2DM, including one-step previous to diabetes and two steps from it with  $OR \geq 1.5$ . The thickness of the arrows is proportional to the OR value. Again, the important role of retinopathy can be visualized. Some chronic conditions could be reached through different paths. Pulmonary heart disease could be directly reached (Supplementary Table S8) from 10 different previous conditions considering the trajectories up to three steps from T2DM and  $OR \geq 1.5$ , including retinopathy, ischemic heart disease, other heart diseases, complicated hypertension, peripheral neuritis, heart failure and atrial fibrillation. Other conditions with a high number of different previous connected diseases were: atrial fibrillation/flutter and other diseases of digestive system (both with 7 different previous diseases connecting them), ischemic heart disease without angina (6 different diseases) and hypertension complicated, heart failure, chronic renal failure, and peripheral neuritis (all with 5 previous different immediate disease connections).

As expected, the number of nodes and significant interactions increased in T2DM networks when we reduced the threshold OR value (Supplementary Table S9). For OR 2.0 there were only 11 nodes and 32 links, for OR 1.5 the network had 23 nodes and 156 links and for OR 1.2 it had 38 nodes and 439 links. For other OR values see Supplementary Table S9. All networks can be visualized in the website.



**Figure 3.** Type 2 diabetes mellitus comorbidity directed network for  $OR \geq 1.5$  (3a, left) and  $OR \geq 2$  (3b, right). The directed T2DM networks considers the temporal associations, filtered to show only chronic conditions with an  $OR \geq 1.5$  or  $\geq 2$  (stronger associations). In the cases where the temporal direction is not clear, the edges are represented with a double arrow (in both directions). The strongest associations from T2DM are to retinopathy, glomerulonephritis/nephrosis, peripheral neuritis/neuropathy and pancreas cancer. T2DM receives the strongest connections from obesity/overweight. Figures created with the MorbiNet application v2.2<sup>25</sup>, <https://morbinet.org/shiny>.

If we only included the patients with more complete data in T2DM networks the number of nodes and interactions were slightly reduced, which can be explained because the number of patients was lower and so was the statistical power. For OR 1.5, when all data were included the network had 23 nodes and 156 interactions and when we included the data with higher quality we had the same number of nodes and 150 links. For other OR values see Supplementary Table S9. Thus, this variable did not seem to have a great impact on the results.

The criterion to define direction of temporal associations in T2DM was arbitrarily set at  $< 40\%$  or  $> 60\%$  of previous/subsequent diagnosis among pairs of chronic conditions. When this criterion was changed to a more stringent threshold of  $< 20\%$  or  $> 80\%$ , there was an important reduction on the number of temporal associations identified. As can be seen in Supplementary Table S10, the number of directional associations were reduced 80% on average, similar for different OR values. This means most directional interactions for pairs of chronic conditions of T2DM networks were between probabilities of 20–40% or 60–80%. In Supplementary table S11 we included the list of directional associations using the  $< 20\%$  or 80% criteria in T2DM network.

## Discussion

We have estimated multimorbidity associations in the general adult population and developed an interactive tool to construct and visualize them graphically, with network parameters. We used a large database, including the information of 3,135,948 adult people with multimorbidity recorded in a long period (12 years), which also allowed estimating temporal direction of the associations. This interactive application can be customized to different needs. Networks can be constructed using different classification systems and can be used both for primary or specialized care. The burden of disease should be considered globally. However, usually in the clinical practice, there is a first chronic disease that occurs, or physicians may be interested in one specific disease because it is more serious or require a more complex treatment. For this reason, it might be useful to extract the comorbidity network for that condition and visualize the diseases associated. This approach might help to analyse the multimorbidity network and to adapt it to clinical practice situations. It is possible to extract subnets for 324 chronic conditions with the ICD-10 system, and 148 chronic diseases using ICPC-2 classification. Apart from the multimorbidity network adjusted by sex and age, stratified networks for gender and different age groups can also be plotted. With this application, both health professionals and the general public can visualize and explore the relationships between diseases according to their specific interests.

We have analysed in detail the T2DM comorbidity network. Most of the conditions T2DM is connected to (Table 2, Fig. 3) are well-known complications of diabetes: retinopathy, renal disease, neuropathy, and cardiovascular disease. T2DM is an important risk factor for cardiovascular events, which is a leading cause of mortality in diabetic patients<sup>26</sup>. Essential hypertension was reported to play an important role in the association between T2DM and comorbid diseases, such as stroke and dyslipidaemia<sup>27</sup> and it is an important risk factor for dying from diabetes<sup>28</sup>.

Retinopathy reflects the damage to microvasculature and is a marker for cerebral vascular disease<sup>29</sup> and system diabetic vasculopathy<sup>30</sup>. Patients with severe stages of retinopathy (macular oedema or proliferative retinopathy)

From	To	Disease connected with type 2 diabetes	OR
T90	F83	Retinopathy	23.8
T90	U88	Glomerulonephritis/nephrosis	3.4
T90	N94	Peripheral neuritis/neuropathy	2.7
T90	D76	Pancreas Ca.	2.4
T90	K75	Acute myocardial infarction	2.1
T90	K92	Atherosclerosis/peripheral vascular dis	2.0
T90	K87	Hypertension complicated	1.7
T90	K82	Pulmonary heart dis.	1.6
T90	K84	Heart dis. other	1.6
T90	F93	Glaucoma	1.5
T90	K91	Cerebrovascular dis.	1.4
T90	T92	Gout	1.3
T90	F84	Macular degeneration	1.2
T83/82	T90	Obesity/overweight	2.6
K74	T90	Ischaemic heart dis. with angina	1.9
X77	T90	Female genital ca, other	1.4
K78	T90	Atrial Fibrillation/flutter	1.4
P72	T90	Schizophrenia	1.3
K89	T90	Transient cerebral ischaemia	1.3
P71	T90	Psychosis organic, other	1.2
<b>Undefined direction</b>			
T90	T89	Diabetes insulin dependent	3.9
T90	K86	Hypertension uncomplicated	2.8
T90	K76	Ischaemic heart dis. without angina	2.2
T90	D97	Liver dis. NOS	1.9
T90	K77	Heart failure	1.9
T90	U99.01	Chronic renal failure	1.8
T90	B82	Anaemia other, unspecified	1.7
T90	T93	Lipid disorder	1.7
T90	K90	Stroke	1.6
T90	D99	Dis. digestive system, other	1.6
T90	B81	Anaemia vitamin B12/folate deficiency	1.4
T90	F92	Cataract	1.4
T90	D98	Cholecystitis/cholelithiasis	1.3
T90	T99	Endocrine/metabolic/nutrition dis. other	1.3
T90	N93	Carpal tunnel syndrome	1.2
T90	D77	Digestive ca. other/NOS	1.2
T90	B78	Hereditary haemolytic anaemia	1.2

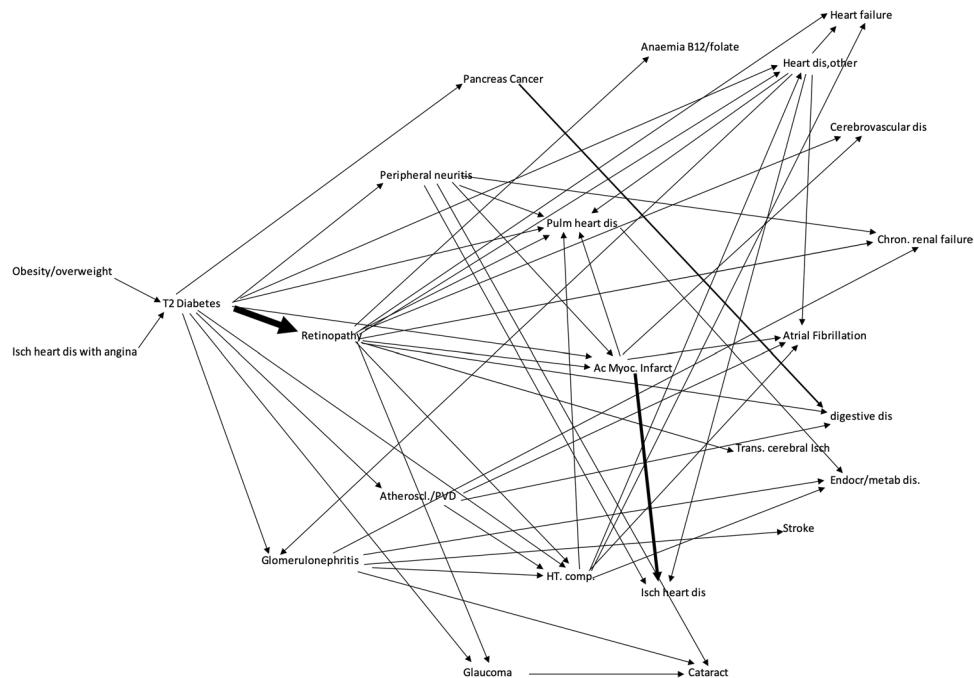
**Table 2.** Odds ratio values for T90 direct links in type 2 diabetes mellitus directed network. T90: Type 2 diabetes mellitus. Abbreviations: Dis: disease, Ca: cancer, NOS: not otherwise specified.

have an increased risk of developing cardiovascular disease<sup>31</sup>. Our results confirm patients with T2DM also have a higher risk of cataracts<sup>32</sup> and open-angle glaucoma<sup>33</sup>.

We found T2DM to be associated with liver diseases, which includes non-alcoholic fatty liver disease. When they both coexist, the risk of complications of diabetes and more severe liver disease increases<sup>34</sup>. Non-alcoholic fatty liver disease is linked with insulin resistance which affects glucose and lipid metabolism, increasing lipogenesis and gluconeogenesis<sup>35</sup>. Diabetes has been independently associated with an increased incidence of anaemia of chronic diseases, which is normochromic and normocytic in the context of inflammatory states<sup>36</sup>. Anaemia is associated with a higher risk of death in patients with T2DM<sup>37</sup>. Atrial fibrillation-flutter also connects to T2DM (Table 2). This association has also been reported in several studies<sup>38</sup>.

Alonso-Moran *et al.* conducted a study in a large data set of patients with T2DM and found more risk in 21 out of 51 chronic diseases<sup>39</sup>. Our OR values are higher, which might be in part because they included older patients. They found an increased risk for emphysema-chronic bronchitis, depression, and psoriasis-eczema, but with OR values were below 1.2. This might be the reason why we did not find these associations. They obtained a reduced risk for malignancies (overall), while we obtained an increased risk for other gynaecological cancers (X77, which includes the uterus, ovary, vagina, vulva, and others), other digestive cancers (D77, which includes mouth, oesophagus, small intestine, liver, biliary cancers and others) and pancreas cancer. T2DM is a known risk factor associated with pancreatic cancer<sup>40</sup>. A higher risk of endometrial cancer has also been observed in patients with T2DM<sup>41</sup>. Insulin resistance, chronic inflammation, and elevated free steroid hormones have been suggested as





**Figure 4.** Temporal comorbid associations. We present the trajectories for T2DM, including one-step previous to diabetes and two steps from it with strong associations ( $OR \geq 1.5$ ). The thickness of the arrows is proportional to the OR value. The important role of retinopathy can be visualized. Abbreviations: dis: disease; Isch: ischemic; T2: type 2; Ac Myoc Infarct: acute myocardial infarction; chron: chronic; Pulm: pulmonary; HT comp: hypertension complicated; Trans: transient; Atheroscl: atherosclerosis; PVD: peripheral vascular disease. Figure elaborated by authors with data exported from the MorbiNet application v2.2<sup>23</sup>, <https://morbinet.org/shiny>.

possible mechanisms involved. Some studies found an increased risk of gastric cancer in diabetes<sup>42</sup>, while others reported little or no change in the risk<sup>43</sup>.

High levels of uric acid were attributed to insulin resistance because of a reduced excretion of uric acid, and recently hyperuricemia has also been reported to have a potential role in incident diabetes<sup>44</sup>. According to our results, T2DM would precede gout more often than otherwise.

We did not find some of the associations reported by Klime *et al.* such as Parkinson or epilepsy, whose study included persons with diabetes who had received inpatient care<sup>45</sup>. Our study was based in general population and we focused on diseases with strong associations, with OR values at least above 1.2, which may hide real associations but with small magnitude.

Regarding directed networks, the analysis showed that both psychosis and schizophrenia precede T2DM (Table 2). Other studies have also found these comorbidities<sup>39,45</sup>. Lifestyle risk factors for diabetes were common in patients with schizophrenia (sedentarism, poor diet, obesity) and atypical neuroleptics have been associated with T2DM<sup>46</sup>.

Carpal tunnel syndrome was associated with T2DM. A meta-analysis suggests that diabetes is a risk factor for carpal tunnel syndrome, and the mechanisms of this increased risk are still being investigated<sup>47</sup>.

These networks show association, and even in the directed networks no causal link should be attributed, since diseases might be associated because they share a risk factor, pathogeny pathway, genetic links or might be caused or influenced by medication (adverse events or protective effect). Many chronic diseases might share some common underlying mechanisms which are driven by oxidative stress<sup>48</sup>.

A prevalent disease can coexist with T2DM, but it can be present no more frequently than in patients without diabetes. The networks we constructed show associations greater than those expected because of their prevalence. Some studies describe the most common comorbidities of T2DM, without considering the prevalence of diseases or having a control group and give complementary information to our results<sup>49–51</sup>.

Networks offer a more global picture because it includes not only direct connections but also indirect associations which give more accurate information about comorbidity. T2DM is not one of the chronic diseases with more connections. In the multimorbidity network, with ICPC-2 nodes and  $OR \geq 1.2$ , the degree is 37 with position 80 out of 148. It is not very central either, its centrality PageRank index is in position 65/148. However, it is connected to nodes with a very high degree and very central in the global multimorbidity network and some of its connections include severe chronic conditions with a poor prognosis. PageRank gives information about the node centrality in the global network. A node with a high value is very connected directly or through other nodes. It connects different clusters or communalities. Pulmonary heart disease has the second-highest PageRank value in the multimorbidity network, although its prevalence is very low. Some conditions with higher prevalence are very central and have a lot of connections in the global multimorbidity network: neuropathy (highest number of connections and 4<sup>th</sup> position in centrality), anaemia (2<sup>nd</sup> in degree and 5<sup>th</sup> more central), and other

digestive diseases (9<sup>th</sup> position both in number of connections and centrality). D99, other digestive system diseases, includes vascular intestine alterations, intestinal malabsorption, celiac disease, pancreas conditions, and others. Cholecystitis/cholelithiasis and cataract have also a high degree (6<sup>th</sup> and 7<sup>th</sup> position out of 148) in the general network. The causal genes of central diseases, with a major impact on multimorbidity, have the potential to influence multiple diseases<sup>52</sup>.

Because of the long study period (2006–2017), we could analyse temporal associations and draw trajectories. They show the relevance of retinopathy in the progress of T2DM. It is not only the condition with the strongest association following T2DM, in Fig. 4 we can see how many chronic diseases have direct links with origin in retinopathy. They include severe conditions with increased disability, reduced quality of life, increased mortality and important use of health resources such as complicated hypertension, cerebrovascular disease, ischemic heart disease, pulmonary heart disease and organ failure (chronic renal failure, heart failure). We have shown the key role of retinopathy. It suggests the importance to prevent it and monitor fundus exploration for early detection. We would like to note in Fig. 4 we just included the conditions with a strong association ( $OR \geq 1.5$ ), otherwise, the picture would be even more complex. Some diseases can be reached from different pathways (Supplementary Table S8), so the risk might be higher than expected if we just looked at immediate connections. Directed networks and disease trajectories consider time and can be useful to identify risk factors and complications of diseases but also study associations to other conditions that take place earlier or later in lifetime and can affect the health status.

Jansen *et al.* studied temporal trajectories of some chronic diseases including diabetes cluster, analysing acute and chronic conditions with hospital encounters data of a large population in 14.9 years<sup>53</sup>. They also identified the retinal disease as a keystone diagnosis marking the progression to worse conditions.

### Limitations

We used SIDIAP database, which contains information from 73% of the Catalan population using the public health care system. Users of private healthcare, presumably with a higher socioeconomic status and probably better health situation were not included. SIDIAP database contains information from electronic medical records. They might include diagnostic errors and some conditions might have been assigned wrong codes or not be recorded in the electronic files. This could conduct to selection bias and limit in part the extrapolation of results, although the large sample size and population-based approach probably protects against major biases.

Directed networks were constructed considering when a diagnosis was made and recorded which does not necessarily imply the moment when the disease started. A condition could have been diagnosed after another one even though it might have started before, leading to wrong temporal assumptions. An additional limitation in T2DM directional networks was the fact that most of the directional interactions were not strongly defined, since 80% of them were in the range of probabilities between 20–40% or 60–80%. That means that only 20% of the temporal associations identified had 80% or more consistency (or less than 20% temporal discordances).

Since the 2014 version, the ICD-10 descriptor for code E11 is “type 2 diabetes mellitus”. But in previous versions, at the time when a lot of the patients might have been diagnosed, the descriptor for E11 was “non-insulin dependent diabetes mellitus” and as some patients with T2DM are treated with insulin this descriptor might create confusion. For this reason, the associations between type 1 and type 2 diabetes should be looked with caution.

In the networks, we included diseases associated to T2DM more than what would be expected because of their prevalence. Thus, frequent diseases can also coexist with diabetes but no more often than in patients without this condition.

### Conclusions

We presented T2DM comorbidities with a strong association and so most of them have already been reported. Our study provides a visual representation based on a large general population and a long follow up period and shows not only direct associations but indirect ones as well as parameters that define the position of T2DM comorbidities in the global multimorbidity network. Neuropathy, anaemia and digestive diseases are comorbidities very connected and central in the general multimorbidity network. We also constructed trajectories that show temporal associations and identify the diseases with an important role in the progression of T2DM, such as retinopathy. Our open interactive website might be useful to explore multimorbidity in a customized way.

### Data availability

The data that support the results of the study are available from SIDIAP. Some restrictions apply to the availability of them. They were released after the signature of a contract and can only be used for the current study, and so they are not publicly available. However, they are available from the authors upon reasonable request and with permission of SIDIAP.

Received: 8 October 2019; Accepted: 10 January 2020;

Published online: 12 February 2020

### References

- Fortin, M., Bravo, G., Hudon, C., Vanasse, A. & Lapointe, L. Prevalence of multimorbidity among adults seen in family practice. *Ann. Fam. Med.* **3**, 223–228, <https://doi.org/10.1370/afm.272> (2005).
- Vos, R., van den Akker, M., Boesten, J., Robertson, C. & Metsemakers, J. Trajectories of multimorbidity: exploring patterns of multimorbidity in patients with more than ten chronic health problems in life course. *BMC Fam. Pract.* **16**, 2, <https://doi.org/10.1186/s12875-014-0213-6> (2015).
- Violán, C. *et al.* Burden of multimorbidity, socioeconomic status and use of health services across stages of life in urban areas: a cross-sectional study. *BMC Public Health* **14**, 530, <https://doi.org/10.1186/1471-2458-14-530> (2014).

4. Williams, J. S. & Egede, L. E. The Association Between Multimorbidity and Quality of Life, Health Status and Functional Disability. *Am. J. Med. Sci.* **352**, 45–52, <https://doi.org/10.1016/j.amjms.2016.03.004> (2016).
5. Panagioti, M. *et al.* Multimorbidity and Patient Safety Incidents in Primary Care: A Systematic Review and Meta-Analysis. *PLoS One* **10**, e0135947, <https://doi.org/10.1371/journal.pone.0135947> (2015).
6. McPhail, S. M. Multimorbidity in chronic disease: impact on health care resources and costs. *Risk Manag. Healthc. Policy* **9**, 143–156, <https://doi.org/10.2147/RMHP.S97248> (2016).
7. Hidalgo, C. A., Blumm, N., Barabási, A. L. & Christakis, N. A. A dynamic network approach for the study of human phenotypes. *PLoS Comput. Biol.* **5**, e1000353, <https://doi.org/10.1371/journal.pcbi.1000353> (2009).
8. Liu, J. *et al.* Comorbidity Analysis According to Sex and Age in Hypertension Patients in China. *Int. J. Med. Sci.* **13**, 99–107, <https://doi.org/10.7150/ijms.13456> (2016).
9. Divo, M. J. *et al.* COPD comorbidities network. *Eur. Respir. J.* **46**, 640–650, <https://doi.org/10.1183/09031936.00171614> (2015).
10. Goekoop, R. & Goekoop, J. G. A network view on psychiatric disorders: network clusters of symptoms as elementary syndromes of psychopathology. *PLoS One* **9**, e112734, <https://doi.org/10.1371/journal.pone.0112734> (2014).
11. Kalgotra, P., Sharda, R. & Croff, J. M. Examining health disparities by gender: A multimorbidity network analysis of electronic medical record. *Int. J. Med. Inf.* **108**, 22–28, <https://doi.org/10.1016/j.ijmedinf.2017.09.014> (2017).
12. International-Diabetes-Federation. (International Diabetes Federation, Brussels, Belgium, 2017).
13. Hermans, M. P. & Dath, N. Prevalence and co-prevalence of comorbidities in Belgian patients with type 2 diabetes mellitus: a transversal, descriptive study. *Acta. Clin. Belg.* **73**, 68–74, <https://doi.org/10.1080/17843286.2017.1348710> (2018).
14. Pouplier, S. *et al.* The development of multimorbidity during 16 years after diagnosis of type 2 diabetes. *J Comorb* **8**, 2235042X18801658, <https://doi.org/10.1177/2235042X18801658> (2018).
15. Wasem, J. *et al.* Co-morbidity but not dysglycaemia reduces quality of life in patients with type-2 diabetes treated with oral mono- or dual combination therapy—an analysis of the DiaRegis registry. *Cardiovasc. Diabetol.* **12**, 47, <https://doi.org/10.1186/1475-2840-12-47> (2013).
16. O'Shea, M., Teeling, M. & Bennett, K. The prevalence and ingredient cost of chronic comorbidity in the Irish elderly population with medication treated type 2 diabetes: a retrospective cross-sectional study using a national pharmacy claims database. *BMC Health Serv. Res.* **13**, 23, <https://doi.org/10.1186/1472-6963-13-23> (2013).
17. Chiang, J. I. *et al.* Associations between multimorbidity, all-cause mortality and glycaemia in people with type 2 diabetes: A systematic review. *PLoS One* **13**, e0209585, <https://doi.org/10.1371/journal.pone.0209585> (2018).
18. Health, W. & Organization. ICD-10 International Statistical Classification of Diseases and Related Health Problems. 10<sup>th</sup> revision. (2016).
19. International, W. & Committee, C. (Oxford University Press, 1998).
20. O'Halloran, J., Miller, G. C. & Britt, H. Defining chronic conditions for primary care with ICPC-2. *Fam. Pract.* **21**, 381–386, <https://doi.org/10.1093/fampra/cmh407> (2004).
21. Zhu, X., Gerstein, M. & Snyder, M. Getting connected: analysis and principles of biological networks. *Genes. Dev.* **21**, 1010–1024, <https://doi.org/10.1101/gad.1528707> (2007).
22. Dong, J. & Horvath, S. Understanding network concepts in modules. *BMC Syst. Biol.* **1**, 24, <https://doi.org/10.1186/1752-0509-1-24> (2007).
23. Devkota, P., Danzi, M. C. & Wuchty, S. Beyond degree and betweenness centrality: Alternative topological measures to predict viral targets. *PLoS One* **13**, e0197595, <https://doi.org/10.1371/journal.pone.0197595> (2018).
24. Freeman, L. Vol. 1, 215–239 (Social Networks, 1978).
25. Aguado, A., Moratalla, F., López-Simarro, F. & Moreno, V. MorbiNet: Multimorbidity Networks in Adult General Population. *MorbiNet v 2.2*. <https://www.morbinet.org> Accessed 1 Oct 2019 (2019).
26. Danaei, G., Lawes, C. M., Vander Hoorn, S., Murray, C. J. & Ezzati, M. Global and regional mortality from ischaemic heart disease and stroke attributable to higher-than-optimum blood glucose concentration: comparative risk assessment. *Lancet* **368**, 1651–1659, [https://doi.org/10.1016/S0140-6736\(06\)9700-6](https://doi.org/10.1016/S0140-6736(06)9700-6) (2006).
27. Kim, H. S., Shin, A. M., Kim, M. K. & Kim, Y. N. Comorbidity study on type 2 diabetes mellitus using data mining. *Korean J. Intern. Med.* **27**, 197–202, <https://doi.org/10.3904/kjim.2012.27.2.197> (2012).
28. Collaboration, G. Bo. M. R. Ff. C. D. Cardiovascular disease, chronic kidney disease, and diabetes mortality burden of cardiometabolic risk factors from 1980 to 2010: a comparative risk assessment. *Lancet Diabetes Endocrinol.* **2**, 634–647, [https://doi.org/10.1016/S2213-8587\(14\)70102-0](https://doi.org/10.1016/S2213-8587(14)70102-0) (2014).
29. Moss, H. E. Retinal Vascular Changes are a Marker for Cerebral Vascular Diseases. *Curr. Neurol. Neurosci. Rep.* **15**, 40, <https://doi.org/10.1007/s11910-015-0561-1> (2015).
30. Moss, S. E., Klein, R., Klein, B. E. & Wong, T. Y. Retinal vascular changes and 20-year incidence of lower extremity amputations in a cohort with diabetes. *Arch. Intern. Med.* **163**, 2505–2510, <https://doi.org/10.1001/archinte.163.20.2505> (2003).
31. Xie, J. *et al.* Association of Diabetic Macular Edema and Proliferative Diabetic Retinopathy With Cardiovascular Disease: A Systematic Review and Meta-analysis. *JAMA Ophthalmol.* **135**, 586–593, <https://doi.org/10.1001/jamaophthalmol.2017.0988> (2017).
32. Li, L., Wan, X. H. & Zhao, G. H. Meta-analysis of the risk of cataract in type 2 diabetes. *BMC Ophthalmol.* **14**, 94, <https://doi.org/10.1186/1471-2415-14-94> (2014).
33. Zhou, M., Wang, W., Huang, W. & Zhang, X. Diabetes mellitus as a risk factor for open-angle glaucoma: a systematic review and meta-analysis. *PLoS One* **9**, e102972, <https://doi.org/10.1371/journal.pone.0102972> (2014).
34. Hazlehurst, J. M., Woods, C., Marjot, T., Cobbold, J. F. & Tomlinson, J. W. Non-alcoholic fatty liver disease and diabetes. *Metab.* **65**, 1096–1108, <https://doi.org/10.1016/j.metabol.2016.01.001> (2016).
35. Samuel, V. T. & Shulman, G. I. Nonalcoholic Fatty Liver Disease as a Nexus of Metabolic and Hepatic Diseases. *Cell Metab.* **27**, 22–41, <https://doi.org/10.1016/j.cmet.2017.08.002> (2018).
36. Lee, Y. G. *et al.* Risk factors for incident anemia of chronic diseases: A cohort study. *PLoS One* **14**, e0216062, <https://doi.org/10.1371/journal.pone.0216062> (2019).
37. Gauci, R., Hunter, M., Bruce, D. G., Davis, W. A. & Davis, T. M. E. Anemia complicating type 2 diabetes: Prevalence, risk factors and prognosis. *J. Diabetes Complications* **31**, 1169–1174, <https://doi.org/10.1016/j.jdiacomp.2017.04.002> (2017).
38. Tadic, M. & Cuspidi, C. Type 2 diabetes mellitus and atrial fibrillation: From mechanisms to clinical practice. *Arch. Cardiovasc. Dis.* **108**, 269–276, <https://doi.org/10.1016/j.acvd.2015.01.009> (2015).
39. Alonso-Morán, E. *et al.* Multimorbidity in people with type 2 diabetes in the Basque Country (Spain): Prevalence, comorbidity clusters and comparison with other chronic patients. *Eur. J. Intern. Med.* **26**, 197–202, <https://doi.org/10.1016/j.ejim.2015.02.005> (2015).
40. Liao, W. C. *et al.* Blood glucose concentration and risk of pancreatic cancer: systematic review and dose-response meta-analysis. *BMJ* **350**, g7371, <https://doi.org/10.1136/bmj.g7371> (2015).
41. Anastasi, E. *et al.* Linking type 2 diabetes and gynecological cancer: an introductory overview. *Clin. Chem. Lab. Med.* **56**, 1413–1425, <https://doi.org/10.1515/cclm-2017-0982> (2018).
42. Yoon, J. M., Son, K. Y., Eom, C. S., Durrance, D. & Park, S. M. Pre-existing diabetes mellitus increases the risk of gastric cancer: a meta-analysis. *World J. Gastroenterol.* **19**, 936–945, <https://doi.org/10.3748/wjg.v19.i6.936> (2013).
43. Miao, Z. F. *et al.* Diabetes mellitus and the risk of gastric cancer: a meta-analysis of cohort studies. *Oncotarget* **8**, 44881–44892, <https://doi.org/10.18632/oncotarget.16487> (2017).

44. King, C. *et al.* Uric Acid as a Cause of the Metabolic Syndrome. *Contrib. Nephrol.* **192**, 88–102, <https://doi.org/10.1159/000484283> (2018).
45. Klimek, P., Kautzky-Willer, A., Chmiel, A., Schiller-Frühwirth, I. & Thurner, S. Quantification of diabetes comorbidity risks across life using nation-wide big claims data. *PLoS Comput. Biol.* **11**, e1004125, <https://doi.org/10.1371/journal.pcbi.1004125> (2015).
46. Suvisaari, J., Keinänen, J., Eskelinen, S. & Mantere, O. Diabetes and Schizophrenia. *Curr. Diab. Rep.* **16**, 16, <https://doi.org/10.1007/s11892-015-0704-4> (2016).
47. Pourmemari, M. H. & Shiri, R. Diabetes as a risk factor for carpal tunnel syndrome: a systematic review and meta-analysis. *Diabet. Med.* **33**, 10–16, <https://doi.org/10.1111/dme.12855> (2016).
48. Barnes, P. J. Mechanisms of development of multimorbidity in the elderly. *Eur. Respir. J.* **45**, 790–806, <https://doi.org/10.1183/09031936.00229714> (2015).
49. Iglay, K. *et al.* Prevalence and co-prevalence of comorbidities among patients with type 2 diabetes mellitus. *Curr. Med. Res. Opin.* **32**, 1243–1252, <https://doi.org/10.1185/03007995.2016.1168291> (2016).
50. Teljeur, C., Smith, S. M., Paul, G., Kelly, A. & O'Dowd, T. Multimorbidity in a cohort of patients with type 2 diabetes. *Eur. J. Gen. Pract.* **19**, 17–22, <https://doi.org/10.3109/13814788.2012.714768> (2013).
51. Bralić Lang, V. & Bergman Marković, B. Prevalence of comorbidity in primary care patients with type 2 diabetes and its association with elevated HbA1c: A cross-sectional study in Croatia. *Scand. J. Prim. Health Care* **34**, 66–72, <https://doi.org/10.3109/02813432.2015.1132886> (2016).
52. Amell, A. *et al.* Disease networks identify specific conditions and pleiotropy influencing multimorbidity in the general population. *Sci. Rep.* **8**, 15970, <https://doi.org/10.1038/s41598-018-34361-3> (2018).
53. Jensen, A. B. *et al.* Temporal disease trajectories condensed from population-wide registry data covering 6.2 million patients. *Nat. Commun.* **5**, 4022, <https://doi.org/10.1038/ncomms5022> (2014).

## Acknowledgements

The study was funded by Instituto de Salud Carlos III, co-funded by FEDER funds –a way to build Europe, grant: PI17/01182 and the Agency for Management of University and Research Grants (AGAUR) of the Catalan Government grant 2017SGR723.

## Author contributions

A.A. and V.M. designed the study. F.M. prepared the database, performed the statistical analysis and constructed the website. V.M. coordinated the analysis and website. A.A., F.M., F.L.S. and V.M. analysed the results. A.A. and F.L.S. wrote and edited the manuscript. All authors provided critical review of manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41598-020-59336-1>.

**Correspondence** and requests for materials should be addressed to A.A. or V.M.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, 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 license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license 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 license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020