Analysis of Comorbidities of Alcohol Use Disorder

Ebrahimi, Ali; Wiil, Uffe Kock; Mansourvar, Marjan; Naemi, Amin; Andersen, Kjeld; Nielsen, Anette Sogaard

Published in:
2021 IEEE Symposium on Computers and Communications (ISCC)

DOI:
10.1109/ISCC53001.2021.9631512

Publication date:
2021

Document version:
Accepted manuscript

Citation for published version (APA):
https://doi.org/10.1109/ISCC53001.2021.9631512

Go to publication entry in University of Southern Denmark's Research Portal

Terms of use
This work is brought to you by the University of Southern Denmark.
Unless otherwise specified it has been shared according to the terms for self-archiving.
If no other license is stated, these terms apply:
• You may download this work for personal use only.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying this open access version
If you believe that this document breaches copyright please contact us providing details and we will investigate your claim.
Please direct all enquiries to puresupport@bib.sdu.dk
Analysis of Comorbidities of Alcohol Use Disorder

Ali Ebrahimi
SDU Health Informatics and Technology, The Maersk Mc-Kinney Moller Institute
University of Southern Denmark
aleb@mmmi.sdu.dk

Uffe Kock wiil
SDU Health Informatics and Technology, The Maersk Mc-Kinney Moller Institute
University of Southern Denmark
ukwiil@mmmi.sdu.dk

Marjan Mansourvar
SDU Health Informatics and Technology, The Maersk Mc-Kinney Moller Institute
University of Southern Denmark
marm@mmmi.sdu.dk

Amin Naemi
SDU Health Informatics and Technology, The Maersk Mc-Kinney Moller Institute
University of Southern Denmark
amin@mmmi.sdu.dk

Kjeld Andersen
Unit of Clinical Alcohol Research, Clinical Institute
University of Southern Denmark
kjeld.andersen@rsyd.dk

Anette Søgaard Nielsen
Unit of Clinical Alcohol Research, Clinical Institute
University of Southern Denmark
ansnielsen@health.sdu.dk

Abstract—Alcohol Use Disorder (AUD) is a clinical diagnosis based on signs and symptoms that are related to excessive alcohol use and it increases the risk of many clinical conditions, psychological instabilities, and social issues. In this study, we aim to identify the comorbidities of Hazardous and Harmful drinkers. We obtained comorbidity networks for Hazardous, and Harmful drinkers using social network analysis techniques. Each network consists of several nodes that represent the diagnostic codes that patients were given during their hospitalization. To precisely identify the most exclusive comorbidities in each drinking group, we proposed a four-step process based on a machine learning algorithm and aggregation functions. Our findings show that the majority of the identified comorbidities in the Harmful drinking group are related to ICD-10 chapters XI, XIX, and XIII. The comorbidities of the Hazardous drinking group, however, did not present a similar clear pattern.

Keywords—Comorbidity, Social network analysis, Machine learning, Alcohol use disorder, Electronic health records

I. INTRODUCTION

Alcohol Use Disorder (AUD) is a clinical diagnosis based on signs and symptoms that are related to excessive alcohol use. It is associated with a variety of physical and psychological diseases such as depression, dementia, liver cirrhosis, upper gastrointestinal cancers, etc. Besides the high prevalence rate of AUD, it is a leading cause of mortality and morbidity [1], with about 3.3 million annual deaths caused by the harmful use of alcohol [2]. Furthermore, this rate is higher in some western countries, like Denmark, where approximately 5.5 percent of all registered deaths per year could be related to alcohol [3]. In many cases, patients and hospital staffs are unaware of the AUD status of the patients. Therefore, relatively few patients receive specialized treatment due to the poor performance of conventional methods of AUD detection and prediction [4].

Electronic health records (EHR) contain data in the form of hospital admission, treatment and discharge, including health information in the form of standardized ICD (International Classification of Diseases) codes [5]. In the remainder of this article, organized as follows. In Section II, the dataset is briefly outlined; the methods employed to develop the network model and identification of comorbidity among patients with drinking problems are elaborated, as well as the ML techniques that are utilized to set the thresholds. The results are presented and discussed in Section III, and finally, Section IV concludes the paper and outlines future work.

II. METHOD

Our methodology consists of three stages, including dataset preparation, network modeling, and comorbidity identification (see Figure 1).
A. Dataset preparation

The dataset for this study was created by linking two sources, the Relay study [11] and EHRs from Odense University Hospital (OUH). The Relay study systematically collected alcohol-related reports from 2,571 patients admitted to the Gastrointestinal, Neurologic and Orthopedic Departments at OUH in the Region of Southern Denmark from October 2013 to June 2016. Patients responded to the Alcohol Use Disorder Identification Test (AUDIT) questioners with a score between 0 and 40. According to Babor, et al. [12], patients can be divided into three drinking groups: scores below 8 are the low-risk group (Normal drinker); scores of 8–15 indicate Hazardous drinker and are the medium-risk group; scores of 16 and higher indicate a Harmful drinker and are the high-risk group. Based on this, patients in our dataset are divided into three groups: 2,114 Normal, 308 Hazardous, and 147 Harmful drinkers. This categorization was used to label the collected EHR from OUH as the target value for training the ML predictive models and for network modelling.

The EHR dataset contains 13,648 administrative records of hospital visits prior to the Relay study for the 2,551 individuals (20 of the patients were dropped from the Relay study due to missing value) who attended the Relay study. The dataset contains historical administrative information in the form of primary diagnoses coded in the Danish version of ICD-10 of patients’ admissions. One of the advantages of the ICD-10 classification is its hierarchical structure that allows us to round the codes to a less specific parent diagnosis chapter or code. We would have faced high dimensionality challenges due to the large number of primary diagnoses. Therefore, based on the hierarchical structure, the ICD codes are rounded to level 3 codes to reduce the number of ICD codes, consequently the number of nodes and edges in the comorbidity network. More information about the dataset can be found in our previous work [4].

B. Comorbidity network modeling

The health trajectory of an individual patient determines the patient’s transition from one disease to another to AUD status during admissions at OUH over time. For example, Patient1 (P1) had Disease 1 (D1), D 2, and D 3 over a period of a year until s/he was found to be a Hazardous drinker as part of the Relay study. Thus, P1’s comorbidity network can be characterized as $D_1 \rightarrow D_2 \rightarrow D_3 \rightarrow AUD_{\text{Hazardous}}$. This is an individual patient’s comorbidity network and it is comprised of nodes and edges. Generally, nodes represent a disease that has appeared in one medical record, whereas the edges are the links between the two nodes. The edge is defined by a weight attribute; that is, the co-occurrence frequency that represents the relationships between the two diseases. The more frequent the two diseases co-occur, the larger the weight of the disease pairs. The formation of the harmful drinker's comorbidity network is depicted in Figure 2.

We created adjacency matrices where each row represents one patient’s disease comorbidity trajectory towards AUD. The created adjacency matrices were then converted into a network that covers all patients. Three different networks are created based on the status of AUD, resulting in the Normal, Hazardous and Harmful drinkers’ networks. This helped us to identify the comorbidities among these three groups. To calculate the centrality and rank of the nodes in the network, four algorithms, degree centrality (DC), closeness centrality (CC), betweenness centrality (BC) and PageRank (PR) were used.

DC is a measure to find a central node based on the number of direct connections to other nodes in the network. It is simply measured by the degree of the node; a higher degree indicates a higher centrality of the node. In this study, the DC identifies diseases that are directly related to other diseases. When a node is easily accessible by other nodes in the network, it is indicated by CC, which measures the distance between a node and other nodes, in particular the shortest path to other nodes in the network. The significance of the disease in the network (as measured by CC) is determined by the average distance from it to every other disease.

Nodes in a network may not be directly connected to each other as there may be more than one intermediate node that bridges the network. BC measures the number of times that a node occurs on the shortest paths (geodesic) that connect any pair of nodes. That is, the higher the frequency of occurrence, the more central the node is. In this study, the topology of BC measures disease complexity and the variety of disease associations. Initially, the use of PR was developed by Brin and Page [13] to assess the importance of website pages. Basically, the number of pages that are linked to a page determines its rank. In this study, the top-ranked nodes are identified through the use of PR that models a random walk of the network as well as based on the number of links possessed by a node relative to others.

C. Identifying comorbidities

As described above, each algorithm has its own perspective and method to measure the centrality of nodes, which in this study is used to indicate how important they are in the AUD comorbidity network. While DC, CC, and BC are classical centrality measures, which determine the

---

**Fig. 1.** Block diagram of research method

**Fig. 2.** Formation of Harmful drinker comorbidity network based on historical EHR data.
centrality of a node based on the structural properties of a network [14], PR is an alpha centrality approach, which has recently been used in many studies to rank the nodes in different application areas [15]. By combining the results of these four measures, we aim to overcome the limitations of each measure and get a more balanced measure of centrality.

We have designed a four-step process, including descending ranking, threshold identification, intersection, and symmetric difference, to identify the most important comorbidities in Hazardous and Harmful drinkers. In this process, first, the diseases are listed in descending order based on their ranks and degrees calculated by DC, CC, BC, and PR, merged and stored in a table. After that, in the second stage, diseases in each drinking group are grouped by the intersection aggregation function to identify the common diseases ranked by the SNA algorithms. For example, Figure 3 displays a thought example using the selected algorithms: DC [D1, D2, D4, D7, D8], CC [D1, D2, D3, D4, D5], BC [D1, D2, D3, D4, D5], PR [D1, D2, D3, D4, D7].

After identifying the common diseases shared in each drinking group, by employing the symmetric difference aggregation function, the comorbidities among the Hazardous and Harmful groups are identified. From the example above, the result of the intersection for Hazardous drinker group is D1, D2, and D4 and for Harmful drinkers, it is D1, D2, and D3. The comorbidity solely for Hazardous drinkers is D4, and for Harmful drinkers it is D3. This helped us identify the most important comorbidities in each drinking group.

In previous studies for the analysis of comorbidities, authors set a threshold to restrict the analysis to a limited number of comorbidities. Setting a threshold facilitates the extraction and representation of the most significant comorbidities among their population. For example, Khan, et al. [7] set a threshold of less than or equal to 30. This was due to the fact that some patients may require frequent admissions for ongoing treatments such as chemotherapy or kidney dialysis. In another study, Shu, et al. [16] restricted their analysis at some point to the top 10 symptoms as well as the top 10 diseases by incidence of the patient, considering that the majority of patient visits to the hospital were attributed primarily to abnormal clinical symptoms. To set the threshold in the current study, we used ML methods to filter the comorbidities for Hazardous and Harmful drinkers.

In this regard, a Random Forest (RF) classifier was trained based on top 10 to 200 nodes (in steps of 10), extracted in the aggregation process, to identify the optimal threshold. Extracted nodes were selected as feature sets, and RF classified them into three classes: Normal drinker, Hazardous drinker, and Harmful drinker. The predictive performance of the constructed models was measured using the predictive accuracy. The values of these performance metrics were computed by using the value of true positive, false positive, false negative, and true negative. The performance of all predictive models was evaluated using 5-fold cross-validation.

III. RESULTS

We constructed three different comorbidity networks for the Normal, Hazardous and Harmful drinker groups. The comorbidities network has 328, 150, and 102 nodes for Normal, Hazardous and Harmful drinkers, respectively. General information regarding the overall properties of each network can be found in Table I.

<table>
<thead>
<tr>
<th>TABLE I.</th>
<th>NETWORK PROPERTIES OF THREE COMORBIDITY NETWORKS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Measures</td>
<td>Normal Drinker</td>
</tr>
<tr>
<td>Number of Nodes</td>
<td>328</td>
</tr>
<tr>
<td>Number of Edges</td>
<td>2,450</td>
</tr>
<tr>
<td>Average edges weight</td>
<td>3.162</td>
</tr>
<tr>
<td>Average Degree Centrality</td>
<td>72.9</td>
</tr>
<tr>
<td>Average Betweenness Centrality</td>
<td>0.045</td>
</tr>
<tr>
<td>Average Closeness Centrality</td>
<td>0.372</td>
</tr>
<tr>
<td>Average Page Rank</td>
<td>0.003</td>
</tr>
</tbody>
</table>

Diseases such as DS72 (Fracture of femur), DZ03 (Medical observation and evaluation for suspected diseases and conditions, ruled out), and DM17 (Gonarthrosis [arthrosis of knee]) are the common diseases among the top five diseases in the comorbidity network of Normal drinkers, ranked by the algorithms. In the Hazardous drinker comorbidity network, DZ03, DR29 (Other symptoms and signs involving the nervous and musculoskeletal systems) and DT93 (Sequelae of injuries of lower limb) are the diseases which are most common among the top five diseases.

Among the Harmful drinkers, DZ03 (Medical observation and evaluation for suspected diseases and conditions, ruled out) and DK70 (Alcoholic liver disease) are the only diseases which are common among the top five diseases in the comorbidity network. DZ03 is a common node among the top 5 nodes in all networks. However, this is increased to five nodes, including DT84 (Mechanical complication of internal joint prosthesis), DM16 (Coxarthrosis [arthrosis of hip]), DZ03, DT93 (Sequelae of injuries of lower limb) and DR29 among the top 10 nodes in all networks.

A. Comorbidities

An RF classifier based on the top 10 to 200 nodes (in steps of 10) shared by the Normal, Hazardous, and Harmful drinker groups was developed to determine the best possible threshold. Figure 4 displays the accuracy of each classifier. The classifier, based on the top 10 nodes, achieved a predictive accuracy of 57%. The predictive accuracy increased greatly by feeding the model with more nodes until the top 70 nodes, which resulted in a predictive

![Fig. 3. Disease comorbidity identification. Degree Centrality (DC), Closeness Centrality (CC), Betweenness Centrality (BC) and PageRank (PR).](image-url)
accuracy of 68%. When the top 130 nodes were used, the accuracy reached a peak of 70%, but then dropped and leveled off in the 69% range. Although the predictive accuracy increased slightly after adding the top 70 nodes, it was steady in the range between 68-70%. Therefore, we decided to set the top 70 nodes as the threshold for further analysis to identify the comorbidities among Hazardous and Harmful drinkers.

Figure 5 displays the result of the symmetric difference process based on the top 70 diagnostic codes among Hazardous and Harmful drinkers. Figure 5 depicts comorbidities among Hazardous drinkers on the left and comorbidities among Harmful drinkers on the right. The diagnostic codes in the middle section are shared by both groups. Our attention is drawn to the asymmetric differences between Hazardous and Harmful drinkers. A total of 17 diseases are identified as comorbidities among Harmful drinkers and 16 diseases are identified as comorbidities among Hazardous drinkers. Furthermore, there are 25 diseases that are shared by both groups.

B. Comparison among comorbidities

Using the method described in the previous section, comorbidities among Hazardous and Harmful drinkers are identified. These diseases are gathered from two networks generated based on SNA techniques. Nodes in these networks are comorbidities and the edges are the co-occurrence of two diseases measured by a weight that increases if pairs frequently co-occur. Four algorithms are used to measure the importance of each node in each group. Table II presents the ranking and degree of 17 comorbidities among Harmful drinkers. Table III shows the degree and ranking of 16 comorbidities among Hazardous drinkers.

Table II shows that DF10 (Mental and behavioural disorders due to use of alcohol) and DI85 (Oesophageal varices) are exclusive comorbidities among Harmful drinkers, and 102 nodes and 306 edges with an average edge weight of 1.696. As shown in Figure 6, only DK74 shares an edge with the other central nodes, such as DF10, DI85, DE87, DK74, and DK86, based on the centrality calculation. The node DF10, which based on Table I, is the most central node among Harmful drinkers, has a relationship with 9 other nodes, which makes it an exclusive node in this manner also. This node, on the other hand, only has a strong edge weight of DK70 (Alcoholic liver disease), which is the most common node group among Hazardous and Harmful drinkers.

In total, the comorbidity network for Harmful drinkers has 102 nodes and 306 edges with an average edge weight of 1.696. As shown in Figure 6, only DK74 shares an edge with the other central nodes, such as DF10, DI85, DE87, DK74, and DK86, based on the centrality calculation. The node DF10, which based on Table I, is the most central node among Harmful drinkers, has a relationship with 9 other nodes, which makes it an exclusive node in this manner also. This node, on the other hand, only has a strong edge weight of DK70 (Alcoholic liver disease), which is the most common node group among Hazardous and Harmful drinkers.

Considering the thickness of edges, there is a strong co-occurrence among diseases related to the digestive system (DK70, DK76, and DK86) from chapter XI-ICD10 and three other chapters, including V-ICD10 (Mental and behavioural disorders), IX-ICD10 (Diseases of the
circulatory system), and XIX-ICD10 (Injury, poisoning and certain other consequences of external causes). As can be seen in Figure 6, this co-occurrence can be seen with the pairs of DF10 and DK70, DI85 and DK76, and DS42 and DK86.

The comorbidity network for Hazardous drinkers has 150 nodes and 580 edges with an average edge weight of 1.856. The pattern in this drinking group is different from the other ones. As shown in Figure 7, the most central nodes, such as DK92, DT92, DM17, and DK21, are connected. It is clear that DK21 shares the edges with DT92 and DM17. Moreover, DK92 and DT92 also share an edge. Looking back at Table III, DT92 and DK21 are two of the most central nodes among Hazardous drinkers.

Considering the weight of the edges, the strongest co-occurrence exists between the pairs DI65 and DI63 and DI65 and DG45 with an average weight of 3.5. DI65 is a disease from chapter IX-ICD10 (Diseases of the circulatory system) that, according to Figure 7, has a co-occurrence with chapter VI (Diseases of the nervous system) due to sharing an edge with DG45 in the hazardous drinkers’ group.

Although there are co-occurrences among chapter XIII-ICD10 (Diseases of the musculoskeletal system and connective tissue) with chapter IX-ICD10 (Diseases of the circulatory system) and chapter XIX-ICD10 (Injury, poisoning and certain other consequences of external causes) through DM17 and DS83, and DM16 and DK21, chapter IX-ICD10 and chapter XIX-ICD10 do share an edge in the comorbidity network among Hazardous drinkers’ group.

IV. CONCLUSION

Comorbidity holds significant medical insights and has its own underlying molecular mechanisms [17], which has been a hot research topic in both the fields of clinical and network medicine [18]. However, most results have been derived from the analysis of clinical data for chronic diseases [7, 8]. This paper investigated the comorbidity patterns in Hazardous and Harmful drinkers among the Danish population. Becoming aware of the high risk of comorbidities of AUD may create incentives for clinicians to detect Hazardous and Harmful drinking and thus be able to advise patients about lowering their alcohol consumption.

### TABLE II. NODES RANK AND CENTRALITY AMONG HARMFUL DRINKERS

<table>
<thead>
<tr>
<th>Rank</th>
<th>Disease</th>
<th>DC</th>
<th>Disease</th>
<th>BC</th>
<th>Disease</th>
<th>CC</th>
<th>Disease</th>
<th>PR</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>DF10</td>
<td>12</td>
<td>DS32</td>
<td>0.036</td>
<td>DE87</td>
<td>0.330</td>
<td>DF10</td>
<td>0.014</td>
</tr>
<tr>
<td>2.</td>
<td>DK86</td>
<td>10</td>
<td>DK86</td>
<td>0.033</td>
<td>DZ00</td>
<td>0.329</td>
<td>DI85</td>
<td>0.014</td>
</tr>
<tr>
<td>3.</td>
<td>DS32</td>
<td>9</td>
<td>DI85</td>
<td>0.019</td>
<td>DF10</td>
<td>0.318</td>
<td>DS32</td>
<td>0.014</td>
</tr>
<tr>
<td>4.</td>
<td>DE87</td>
<td>9</td>
<td>DF10</td>
<td>0.016</td>
<td>DI85</td>
<td>0.317</td>
<td>DK86</td>
<td>0.013</td>
</tr>
<tr>
<td>5.</td>
<td>DK74</td>
<td>9</td>
<td>DS91</td>
<td>0.015</td>
<td>DK86</td>
<td>0.315</td>
<td>DK74</td>
<td>0.009</td>
</tr>
<tr>
<td>6.</td>
<td>DK86</td>
<td>8</td>
<td>DR10</td>
<td>0.015</td>
<td>DS81</td>
<td>0.302</td>
<td>DL08</td>
<td>0.009</td>
</tr>
<tr>
<td>7.</td>
<td>DG62</td>
<td>8</td>
<td>DZ00</td>
<td>0.014</td>
<td>DK74</td>
<td>0.306</td>
<td>DG62</td>
<td>0.009</td>
</tr>
<tr>
<td>8.</td>
<td>DS81</td>
<td>8</td>
<td>DE87</td>
<td>0.011</td>
<td>DS82</td>
<td>0.293</td>
<td>DE87</td>
<td>0.009</td>
</tr>
<tr>
<td>9.</td>
<td>DZ00</td>
<td>7</td>
<td>DG62</td>
<td>0.009</td>
<td>DQ65</td>
<td>0.277</td>
<td>DS81</td>
<td>0.008</td>
</tr>
<tr>
<td>10.</td>
<td>DM25</td>
<td>6</td>
<td>DQ65</td>
<td>0.009</td>
<td>DQ62</td>
<td>0.275</td>
<td>DR10</td>
<td>0.008</td>
</tr>
<tr>
<td>11.</td>
<td>DL08</td>
<td>6</td>
<td>DM25</td>
<td>0.008</td>
<td>DM25</td>
<td>0.273</td>
<td>DM25</td>
<td>0.008</td>
</tr>
<tr>
<td>12.</td>
<td>DQ65</td>
<td>5</td>
<td>DR74</td>
<td>0.007</td>
<td>DR74</td>
<td>0.272</td>
<td>DS91</td>
<td>0.008</td>
</tr>
<tr>
<td>13.</td>
<td>DR74</td>
<td>5</td>
<td>DS81</td>
<td>0.006</td>
<td>DL08</td>
<td>0.272</td>
<td>DZ00</td>
<td>0.008</td>
</tr>
<tr>
<td>14.</td>
<td>DS51</td>
<td>5</td>
<td>DK74</td>
<td>0.004</td>
<td>DS51</td>
<td>0.246</td>
<td>DS70</td>
<td>0.007</td>
</tr>
<tr>
<td>15.</td>
<td>DS91</td>
<td>4</td>
<td>DL08</td>
<td>0.001</td>
<td>DS70</td>
<td>0.244</td>
<td>DQ65</td>
<td>0.007</td>
</tr>
<tr>
<td>16.</td>
<td>DR10</td>
<td>4</td>
<td>DS70</td>
<td>0.0007</td>
<td>DS91</td>
<td>0.239</td>
<td>DR74</td>
<td>0.007</td>
</tr>
<tr>
<td>17.</td>
<td>DS70</td>
<td>4</td>
<td>DS51</td>
<td>0.0007</td>
<td>DR10</td>
<td>0.226</td>
<td>DS51</td>
<td>0.006</td>
</tr>
</tbody>
</table>

Fig. 6. Comorbidity network among Harmful drinkers.

Fig. 7. Comorbidity network among Hazardous drinkers.
The present study was carried out on 351 diseases among 2,114 Normal drinkers, 308 Hazardous drinkers, and 147 Harmful drinkers in a full range of age groups (18-110) from the Region of Southern Denmark. Our population consisted of patients from Gastrointestinal, Neurologic, and Orthopedic departments at OUH. We focused on the development of a comorbidity network based on SNA techniques and algorithms to identify the most important comorbidities among Hazardous and Harmful drinkers. Furthermore, we investigated the relationships between the characteristics of the comorbidity network among each drinking group.

We have demonstrated that it is possible to use EHRs to identify disease occurrence and detect the risk factors of comorbid conditions for people with AUD. Our comorbidity networks offer insights that can help clinical staffs to better understand high-risk diseases and progression patterns among Hazardous and Harmful drinkers. Such a network representation can be an effective way to visualize the health journey of a specific patient with AUD.

Despite our best efforts, this study has some limitations. The first limitation is that the data was collected from a small population who were admitted to only three departments, including Gastrointestinal, Neurologic, and Orthopedic, at the OUH. Therefore, the health trajectory consisted of disease codes that are likely to be serious enough to occur during hospitalization in those departments. Having a national Danish dataset may give us more insight about diagnosis codes that were missing in the current dataset or have a low frequency that would otherwise be included or have a high frequency if more data were available. Another limitation is the lack of considerations on the temporal aspect of AUD progression (disease trajectory toward AUD). Both limitations will be addressed in our future work.

In summary, the main contribution of this study was the development of comorbidity networks for Hazardous, and Harmful drinkers to identify the most exclusive comorbidities for each group. This has been done by proposing a three-stage methodology, including Data Preparation, Network Modeling, and Comorbidity Identification. The heart of the methodology, comorbidity identification, consisted of a four-step process which after storing extracted comorbidities in descending order, an RF classifier was employed to identify the best possible threshold for limiting the number of commodities, thus excluding insignificant ones. After that, by the means of intersection and symmetric differences techniques, the most exclusive comorbidities for each drinking group were identified.

### Reference


