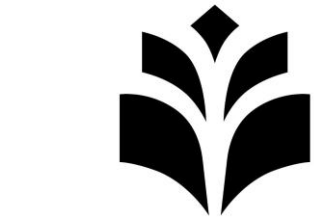

UNIVERSITY OF
EASTERN FINLAND

Sequential Pattern Mining for Analyzing Disease Progression from Patient Health Data

Sazid Nur Ratul

Master's Thesis

School of Computing
Computer Science

January 2026

Abstract

In medical research, one of the primary objectives is to identify disease associations to enable early treatment and improve patient outcomes. The objective of this study was to evaluate the feasibility of achieving this goal by identifying disease association patterns in the longitudinal medical records of more than 3.9 million Finnish patients. We utilized the FP-Growth algorithm to find the frequent disease sequence associated with each patient's timeline and then used Relative Risk, 95% Confidence Intervals ,and Relative Width to establish the clinical validity of these identified patterns.

We completed our analyses at two different levels of ICD-10 classification: specifically, single code level (e.g., E10) and block level (e.g., E10-E14). For the single code level of classification, we found patterns for 1,473 single codes, and for the block level, we found patterns for 206 categories. Our results demonstrated several clinically relevant patterns in association with multiple diseases. Among notable findings, diabetic retinopathy disease patterns clearly demonstrated associations between diabetes (E10, E11) and retinal disorders (H36). Cardiorenal disease patterns were also demonstrated for acute kidney failure (N17), with heart failure combination diseases resulting in relative risks above 10. Diabetic foot disease patterns were clearly demonstrated through osteomyelitis (M86), where diabetes and lower limb ulcers resulted in relative risks of greater than 150. The multi-organ nature of amyloidosis (E85) was also demonstrated through the identification of associations between amyloidosis and cerebral amyloid angiopathy, renal disease, and cardiac disease. Finally, drug poisoning (T36-T50) disease patterns demonstrated clustering among mental health disorders and alcohol abuse, consistent with self-harm presentations.

At the block level, we identified larger disease associations. Among important findings, lung diseases due to external agents were associated with both pleural disease and aspiration-prone conditions in neurological patients. Expected vascular risk factor clustering and post-stroke cognitive sequelae were demonstrated for cerebrovascular diseases. Similarly, renal failure at the block level reinforced the cardiorenal syndrome disease patterns demonstrated at the single code level.

We were able to successfully identify high-risk combinations of disease with reliable statistical evidence. Our narrow confidence intervals, along with low relative widths, provide reliable risk estimates that could be used in clinical considerations. This will enable us to use this data to enhance predictive health care by identifying prior conditions that have an increased likelihood of producing a variety of diseases, thus enabling the possibility of early detection and prevention of disease progression. The result from this study is publicly available for research purposes here: **https://cs.uef.fi/ml/impro/disease-pattern**

Keywords: Disease Progression, Medical Data Analysis, Sequential Pattern Mining, FP-Growth, Relative Risk, Confidence Interval, Relative Width

Acknowledgments

I am grateful to Almighty God who gave me all the knowledge and the path that led me to my current stage of life. I would like to express my sincere gratitude to my supervisor, Professor Pasi Fränti. It's been such a pleasure and honor to work under his supervision. I sincerely appreciate Jimi Tuononen for helping me from time to time whenever I needed. I am also grateful to the School of Computing, the Faculty of Science, Forestry and Technology, and the University of Eastern Finland for their support and resources.

I would like to thank my beloved wife for always pushing me to finish my work. Her helping mindset always motivated me. And lastly, to my parents, who have always made my education their top priority and helped me get this far.

Table of Contents

List of abbreviations

RR	Relative Risk
CI	Confidence Interval
RW	Relative Width
ICD-10	International Classification of Diseases, 10th Revision
SE	Standard Error
CKD	Chronic Kidney Disease
AKI	Acute Kidney Injury

List of Tables

List of Figures

1. Introduction

The growing number of chronic disease patients worldwide creates a large burden on global health care systems. Chronic health issues are the number one cause of death in the world right now. In 2002, major chronic diseases caused 29 million deaths [1]; by 2021, that number had grown to 43 million [2]. As per the World Health Organization (WHO) [2], Cardiovascular disease was the biggest cause of death in 2021, with a death toll of 19 million. The list includes cancer, respiratory disease, and diabetes. Cardiovascular disease is also responsible for 80% of premature deaths. The WHO also says that [3] noncommunicable diseases (NCDs), often known as chronic diseases, cause 41 million deaths per year around the world. These diseases include diabetes, cancers, and heart and lung conditions. What is most concerning is that 17 million of those deaths were people under 70. The situation is even worse in developing countries, where 82% of premature deaths are related to chronic conditions. These countries are facing socioeconomic challenges related to the deaths of working-age people [3].

Woolf and Schoomaker[4] recently conducted an analysis of death rates in the U.S. from 1959 to 2017. Their findings indicate that more than 2.2 million working-age adults died due to chronic disease in the United States in 2017; specifically, chronic liver and cardiovascular disease contributed to many of the deaths. Additionally, they found that the death rate for working-age Americans (ages 25-64) increased substantially from 2010 through 2017. Woolf and Schoomaker also determined that the working-age death rate due to liver disease alone increased by approximately 40.6% from 2010 through 2017. While the mortality rate of chronic disease is a serious concern, chronic conditions also have a considerable social and economic burden on both individuals and healthcare systems.

As a result of the increasing prevalence of disease complications, increasing death rates among working-age people, and changing lifestyles and diets, it has become increasingly important to understand how disease develops and how different diseases relate to one another so that early intervention may be possible and better patient outcomes may be achieved.

In order to identify how diseases influence one another and correlate, we implemented the FP-Growth algorithm to mine patterns from the medical records of 3.9 million Finnish people. We also used statistical methods like Relative Risk to evaluate the strength of the mined patterns.

Additionally, Confidence Interval and Relative Width metrics were calculated to see the reliability of the relative risk and to get a better idea of how accurate the predictions were. This research aims to identify disease patterns with a high likelihood of recurrence, potentially helping doctors in decision-making and contribute the field of predictive healthcare.

2. Related Work

In recent years, the medical field has been well-adapted to utilizing technology for managing and storing medical records. For this reason, medical research has shifted towards using patient health data to predict disease development. In 2023, Choudhary and Fränti presented a method for predicting how diseases develop in patients based on past disease history. They analyzed a large dataset of 3.9 million patient records and turned those health records into simple but large networks where each disease is a node, and each edge represents how the disease is connected to others. By using the supervised depth-first search technique, they found common patterns of disease occurrences that can help predict future health risks. They showed a comparison table of disease occurrences using forward, backward, and individual link frequencies with conditional probabilities and relative risk values, where relative RR > 1 means a higher probability of disease progression. Their method achieved an AUC score of 0.68 and an F1 score of 0.13 for predicting disease sets. They also mentioned issues with current methods, such as the difficulty of tracking diseases over long periods. They address these problems and provide an approach to identifying disease development based on age and sex. As such, this research is relevant to our work since we will be using similar patient data to identify patterns of disease and predict potential health risk factors.

In another research, Uddin et al. (2022) [5] identified several chronic disease development relationships and presented the fact that one chronic disease can greatly affect many others. They studied 29,000 patients for 24 years, 1995-2018. Using the t-test statistical method, their study was statistically significant when comparing the progression of disease between chronic health conditions. The results indicate that there is a considerable amount of interaction between certain chronic diseases, such as cardiovascular disease and diabetes, and the progression of those interactions can vary depending on the age and/or sex of the individual. In addition, Su et al. (2016) [6] found a connection between asthma and a wide range of comorbidities and proposed that chronic inflammatory conditions can cause increased risk of developing cardiovascular disease through systemic inflammation. In 2020, Rocca et al. argue that treatment for specific chronic conditions is not enough. They showed that multiple chronic diseases can coexist, making treatment and patient outcomes very difficult [7]. In another 2020 cohort study, Elewonibi and Nkwonta suggested that chronic health issues might impact health behaviors and indirectly increase the severity of other diseases [8] Therefore, these studies highlight the importance of understanding disease interconnectivity. Our research expands this

domain by analyzing disease trends and predicting potential health threats at the population level.

3. Research Methodology

The purpose of this section is to provide an overview of the methodology used in this study. It describes the methodical processes for gathering, processing, analyzing, and interpreting data. The following flow chart shows our steps.

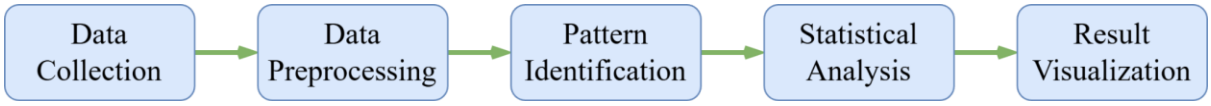

Figure 1: Research methodology

Each step is explained briefly in the following sections.

3.1. Data Collection

In this study, the dataset was provided by the Finnish Institute for Health and Welfare (THL, Terveyden ja hyvinvoinnin laitos) with 3,987,382 patient records. The database contained 9,041,017 entries in total. Each entry is a single diagnostic or medical event for a specific patient with an ICD-10 code. The database has the following variables:

- Pid: Patient identification number in the dataset.
- ICD-10: Disease classifications according to the International Classification of Diseases, 10th revision.
- Sequence: The amount of time (in minutes) that elapsed from the moment the patient entered the system until this disease was documented. Sequence = 0 means it is the patient's first recorded illness, and Sequence = 20,000 means the disease was documented 20,000 minutes after they first entered the system.
- Level: Specifies the code detail level:

- Level 1: Value “1” indicates a specific illness/disease diagnosis. i.e E10, E11, E12 etc.
- Level 2: Value “2” indicates a broad classification group that includes similar illnesses/diseases within a single group or block. For example, with an ICD-10 code E10 and level “2”, the entire range of “E10-E14” includes all forms of diabetes mellitus.

3.2. Data Preprocessing

Before we can mine for patterns in data, we first have to pre-process our data to remove several irrelevant diseases from the dataset and also to separate the two different types of data contained within this dataset (level one and level two) and prepare the raw data into a form suitable for use during the mining process. In order to identify meaningful disease patterns, we deleted every record that had a disease code that began with the letters V, W, X, Y, Z, R, or U, as they represent external causes of death, or other variables that contribute to an individual's health status, or symptoms, and therefore do not represent a real disease.

After removing unnecessary data, the dataset is still not ready for mining patterns. The data consists of individual records for each disease event per patient, but most mining algorithms use all events for a single identifier as one transaction. This means we need to create a single record containing all diseases experienced by each patient, sorted chronologically. We also have to handle level 1 and level 2 records separately when constructing these single records. For example, consider a patient with the following records:

icd10	level	seq
R55	1	0
J35	1	391620
I83	1	1461540
R50	2	0

J30	2	391620
I80	2	1461540

Table 1: Example Dataset

So, the constructed record will appear as follows:

level 1	level 2
R55 > J35 > I83	R50 > J30 > I80

Table 2: Constructed record for example dataset

Both records are sorted according to their sequence. For Level 1, the patient's first disease is R55 because the sequence is 0. After 391,620 minutes (~ 272 days), the patient developed J35. Similarly, I83 was recorded 1,461,540 minutes (~1,015 days) after the initial entry (R55).

In this example, the sequence for both levels is identical, but the ICD-10 code is not. As explained earlier, level 1 represents a specific disease code, and level 2 represents a broader disease group. Our dataset contains separate records for the same disease event at each level for a patient. For example, I80 at Level 2 represents the entire group of diseases coded from I80-I89. Therefore, I83 in Level 1 becomes I80 in Level 2.

Our study needs to calculate patterns for all the unique diseases. For a specific target disease, having all patient timelines or complete disease records from a patient is unnecessary. This means we need to create transactions that are meaningful to pattern mining. We have some constraints in calculating patterns for a disease:

1. To be included in the mining process for a specific target disease, the patient must have the target disease in the timeline.

2. The only diseases we are concerned with are those that occur before the target disease, and all other diseases from a patient's timeline (including the target disease itself) will be removed for this calculation.

These constraints help the analysis to produce only meaningful and feasible patterns and ensure that the research is more coherent. As an example, if the target disease is D80:

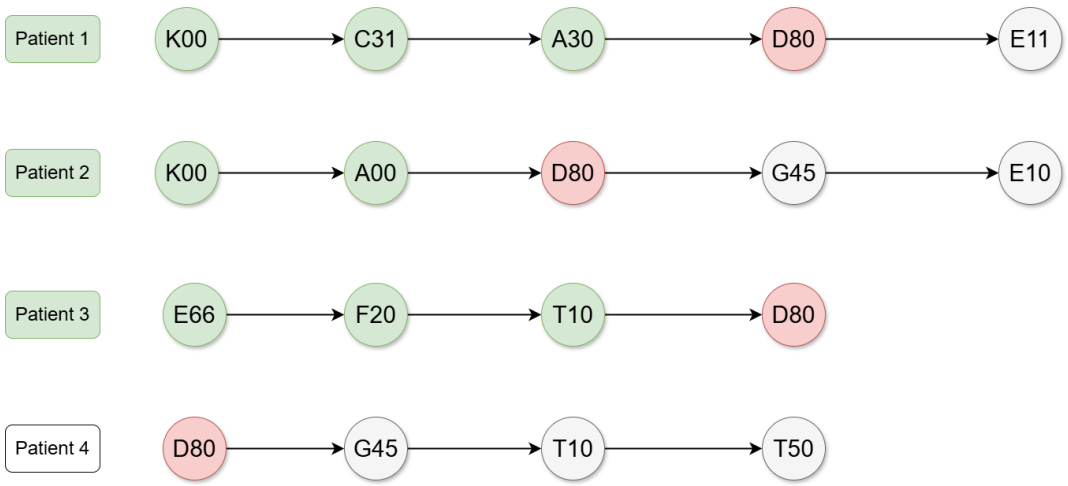

Figure 2: Example patient timeline for target disease D80

In this figure,

- All of the patients have target disease D80 in their timeline.
- Patients 1 to 3 have valid timelines for transactions because they have diseases (green nodes) recorded before the first occurrence of D80 (red node) in their timelines.
- Patient 4, however, is not a valid transaction since D80 appears in the very first position of their timeline, meaning no preceding diseases exist.

So the final transactions for this example D80 are:

Transactions
K00 > C31 > A30
K00 > A00

E66 > F20 > T10

Table 3: Example transaction for target disease D80

After removing unnecessary data, making single transactions, and reconstructing transactions for all unique diseases, our data preprocessing part is complete. These transactions are ready for mining patterns. After preprocessing, we found 1473 unique diseases with level 1 and 206 unique diseases with level 2. The disease count with level 1 is expected to be higher than level 2 because level 1 is more specific and represents a single disease.

3.3. Pattern Identification

We used several techniques and algorithms with our preprocessed data for mining patterns. We started with the brute force technique and the Depth-first search (DFS) algorithm, and later utilized Apriori, Eclat, and FP-Growth algorithms. Each of them has advantages and disadvantages. The comparisons are shown in the table below:

Technique and Algorithms	Comments
Brute-force and DFS	This approach is suitable for finding patterns with only a single disease, but it requires a lot of resources to find patterns with multiple diseases because our dataset is huge. Since the brute-force method involves manual calculations, there is a possibility of creating erroneous patterns, and the likelihood of this error occurring is much higher when compared to other methods.
Apriori	Apriori produces the types of patterns we want; it produces both patterns involving a single disease and patterns involving multiple diseases. However, like the brute force method, Apriori requires a tremendous amount of processing power to operate on our large data set and takes days to complete.

Eclat	A major limitation of Eclat is that it only identifies patterns involving multiple disease combinations. However, we are also interested in patterns that only contain a single disease, and a single disease can be sufficient to impact other disease development. But the Eclat consumes fewer resources compared to Apriori.
FP-Growth	The FP-Growth algorithm consumes fewer resources than others, providing patterns with single and multiple diseases. We were able to produce patterns in a reasonable time. This algorithm fulfills our needs in every way.

Table 4: Pattern mining techniques and comparison

To ensure accuracy, we tested all algorithms and approaches on a smaller subset of our dataset and confirmed that they generated identical patterns with the same results.

3.3.1. FP-growth Algorithm

Frequent Pattern Growth (FP-Growth) is an advanced and efficient algorithm for frequent pattern mining, particularly suited for large datasets [9]. FP-Growth constructs a compact data structure called an FP-tree, which allows for the fast identification of frequent patterns without needing candidate generation. Candidate Generation refers to the repeated creation of item-set combinations (candidates), which are checked for their frequency in the database. The most common form of candidate generation is through the use of algorithms such as Apriori, where single items are first created as candidates; these are scanned across the entire database to determine their frequency. Once the single items have been determined to be frequent, additional items are added to create new candidates. Again, these are scanned across the entire database to check their frequency. By building the FP-Tree, FP-Growth eliminates the need to generate and test candidates repeatedly, and therefore can operate at a significant speed advantage over candidate-generating algorithms. In summary, the reasons behind choosing FP-Growth are:

- Faster than others due to no need for candidate generation.
- It is memory-efficient because it uses a tree-based structure to store transaction data compactly.
- Can efficiently mine disease patterns from large patient records.

3.3.2. FP-Growth Methodology

FP-Growth works by first constructing a compact data structure called the FP-tree from the dataset and then recursively extracting frequent patterns. The process starts with scanning all transactions for the target disease to count item frequencies, generate item sets, and create an FP tree by inserting all items. The algorithm requires the support parameter for mining. FP-Growth uses the support threshold when selecting items for the FP tree and identifying the final frequent patterns. This support parameter is crucial because a large dataset can produce a huge amount of patterns by generating all possible combinations from the transactions. For support 1, the algorithm will include all the combinations of patterns that occurred together only once in any patient's disease record. These single-frequency patterns are not a helpful or interesting case in our study. Also, lower support needs higher computing resources to find, store, and analyze the results. For our research, we set a support value of 5% for all target diseases with a minimum of 5 patients. We also tried several supports, such as 2, 3, and 10. Higher support misses some patterns that might be useful in our results, and lower support produces unnecessary results with high resource consumption. The algorithm will remove any pattern that doesn't meet the minimum support threshold. The remaining valid patterns will be sorted in descending order of frequency to maintain consistency.

The tree is constructed by inserting items as nodes while sharing common prefixes with existing paths. There will be a new node if no existing path matches. Each node in the tree stores a count and pointers to utilize later for traversal. After building the tree, frequent pattern mining is performed by recursively traversing the tree and extracting conditional patterns. The process continues until no frequent patterns remain.

3.3.3. Example of Pattern Mining with FP-Growth

This section will demonstrate the process of mining patterns using the FP-Growth algorithm by a specific example. We will focus on the target disease, N17, and analyze the medical histories of six patients.

The medical histories of the patients are as follows:

Patient	Disease Timeline
P1	N00 > F30 > F20 > D60 > I30 > N17
P2	I70 > N10 > N00 > D60 > N17 > P30 > M17
P3	D60 > I70 > N10 > N17 > E10
P4	I30 > D60 > N00 > I70 > N17
P5	A00 > M00 > M15 > E00 > E11
P6	N17 > M17 > E10 > P30

Table 5: Patient timeline for pattern mining example

Now, we have to process the data. From these medical histories, we can see that Patient 5 does not have the target disease N17, so we will exclude this patient from our analysis. Patient 6 has N17 listed first in their medical history, but since no diseases were recorded before N17, this patient's data is not helpful for our analysis. Therefore, we will only consider the diseases that occurred before the diagnosis of N17 in the other patients' histories.

After processing the data, we create transactions for the FP-Growth algorithm based on the relevant diseases:

ID	Transaction
T1	N00 > F30 > F20 > D60 > I30
T2	I70 > N10 > N00 > D60
T3	D60 > I70 > N10

T4	I30 > D60 > N00 > I70

Table 6: Generated transaction for pattern mining example

Next, the FP-Growth algorithm counts the frequency of each item in the dataset and sorts them based on their counts for global order. If two items have the same count, they are sorted alphabetically. For this example, the algorithm will ignore items not meeting our minimum support threshold set at 2. The sorted frequency counts for the items are as follows:

Item	Frequency
D60	4
I70	3
N00	3
I30	2
N10	2

Table 7: Frequency count table for pattern mining example

From these counts, the algorithm will establish a global order of the items based on their frequency: D60 > I70 > N00 > I30 > N10

Using this global order, FP-Growth will reorder the transactions internally as follows:

ID	Transaction
T1	D60 > N00 > I30
T2	D60 > I70 > N00 > N10
T3	D60 > I70 > N10
T4	D60 > I70 > N00 > I30

Table 8: Re-ordered transaction for pattern mining example

With the reordered transactions, the FP-Growth algorithm constructs an FP-tree. This tree is structured from top to bottom, but can traverse from bottom to top during the counting process or reach related nodes in any order. The FP-Tree for the above transactional data is shown below:

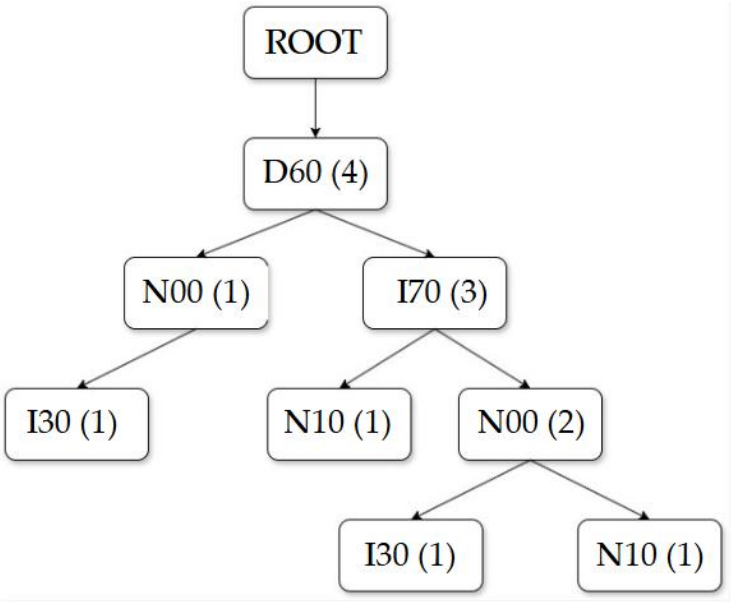

Figure 3: FP-Tree for pattern mining example

Once the FP-Tree has been constructed, we can find the frequent patterns by recursively counting the frequency of each pattern. As a result of the FP-Growth Algorithm, we eliminate patterns with less than the specified minimum support threshold of 2. The identified frequent patterns for this example follow:

Patterns	Count
D60	4
I70	3
N00	3
I30	2
N10	2
D60, I70	3
D60, N00	3

D60, I30	2
D60, N10	2
I70, N00	2
I70, N10	2
N00, I30	2
D60, I70, N00	2
D60, I70, N10	2
D60, N00, I30	2

Table 9: Generated pattern for example disease N17

The generated patterns contain single and multiple items, and all other patterns that don't meet support requirements are removed. This example includes D60, I70, and N00 as the three most common diseases; they appear 4, 3, and 3 times in the patient's timelines prior to N17. The other remaining patterns continue with a count of 2. In addition, when a pattern contains more than one disease, it implies that all diseases within this pattern occur together or individually before the target disease occurs.

The following sections describe the statistical methods used to determine how important the mined patterns are. As we said earlier, the frequency of patterns alone does not indicate their analytical significance; so, other measures were applied to evaluate their statistical reliability.

3.4. Statistical Analysis

We evaluated the strength and statistical significance of the discovered patterns using relative risk (RR), confidence intervals for RR, and the relative width of these intervals. The RR indicates how many times more likely a person with a specific pattern is to get the target disease than a person without that pattern. For example, if RR=2.5, it indicates that individuals with the pattern are 2.5 times as likely to develop the disease as individuals with no pattern. As a result, we only examine RR >1 since RR <1 would indicate that there is either no association between the patterns and the target disease or a lower association.

To determine whether the findings were statistically significant, we used a 95% CI for each RR. A Confidence Interval (CI) gives us the upper and lower limits for the RR , and this helps us understand where the true RR is most likely fall. So, if the CI for an RR includes only values greater than 1, we can conclude that the observed RR is statistically significant with a strong relationship to the target disease. Alternatively, if the CI includes only values less than 1, then we can conclude that there is a statistically significant negative association between the risk factor(s) and the target disease, therefore the association is relatively unimportant.

We also evaluated how reliable each RR was by calculating the Relative Width (RW) of its confidence interval. RW shows how precise or dependable the RR estimate is. We defined two thresholds for RW: values of 0.5 or less indicate that the RR is very reliable, values between 0.5 and 1 suggest moderate reliability, and values above 1 indicate low reliability, meaning the RR is not very useful. The RW is the width of the confidence interval divided by the RR. In other words, an RW of 0.5 means that the CI's width is half of the RR's value, and a RW of 1 means the CI's width is the same as the RR's value. The reason for these thresholds is to categorize risk estimates by their accuracy. If the RR is 2.5 with a 95% CI of [2.3, 2.8], the RW is (2.8 – 2.3)/2.5 = 0.2. This narrow range shows that the uncertainty is small, and the values are well above 1. This indicates the pattern has a high disease risk, is statistically significant, and very reliable. Similarly, if the RR is 2.5 with a 95% CI of [1.2, 5.70] (RW ≈ 1.8), the interval is very wide and bigger than the RR itself. This means there is a lot of uncertainty, patterns are not reliable, and the RR is not precise.

3.4.1. Relative Risk (RR)

Relative Risk (RR) is commonly used in cohort studies to compare the risk of a particular event between exposed and unexposed groups. In our study, the exposed group includes patients with a specific disease pattern, while the unexposed group includes those without it. The event of interest is the occurrence of a target disease. We use RR as a statistical measure in our study because it clearly shows how much more or less likely the target disease is to occur in patients with the disease pattern compared to those without it. Additionally, Zhang and Yu (1998) [10] discussed that when outcomes are common, using RR provides a more accurate measure of

association than using the odds ratio method, and the odds ratio can misinterpret risk. The equation below calculates RR by comparing the incidence rates of the target disease in both groups [11].

$$RR = \frac{a}{a + b} / \frac{c}{c + d}$$

Here:

- a = Patients with the disease pattern and target disease.
- b = Patients with the disease pattern but without the target disease.
- c = Patients without the disease pattern but with the target disease.
- d = Patients without the disease pattern and target disease.

Thus,

- a + b = Total patients with disease pattern
- c + d = Total patients without disease pattern

Table X is a 2×2 contingency table that shows the counts for patients with or without the disease pattern, and with or without the target disease. In epidemiological studies, this table is often used to find out RR measurements.

	Target Disease Present	Target Disease Absent	Total
Disease Pattern Present	a	b	a+b
Disease Pattern Absent	c	d	c+d
Total	a+c	b+d	a+b+c+d

Table 10: 2×2 Contingency Table of Disease Pattern and Target Disease

This table supports the RR calculation by showing the difference in incidence rates between the exposed group (those with the disease pattern) and the unexposed group (those without the disease pattern).

3.4.2. Confidence Intervals (CI) for RR

The Confidence Interval provides us with an approximate range for the true RR. If the CI is narrow, the findings can be considered as being more precise; if the CI is wide, then the findings can be viewed as being less certain. Typically, a 95% CI is used in cohort studies. This means there is a 95% chance that the true RR falls within this range. The CI shows how reliable the calculated RR is when comparing incidence rates.

The reason we chose CI over other methods is that it not only provides an estimate but also shows how uncertain that estimate is. This measure of uncertainty is crucial when interpreting a large dataset like ours, as it helps prevent misleading conclusions that could arise from looking at RR alone. In 1978, a study by Katz et al. [12] showed how CI is a better way to understand the precision of the risk ratio in cohort studies, particularly when working with large patient datasets [13]. We used Wald's method to calculate the CI because it is simple and widely used in epidemiological studies [12, 14]. Wald's method uses the natural logarithm (ln) of the RR and its standard error. The equations for calculating the CI are as follows:

Standard Error:

$$SE = \sqrt{\left(\frac{1}{a} - \frac{1}{a+b} + \frac{1}{c} - \frac{1}{c+d}\right)}$$

Here, a, b, c, and d represent the same definitions as in Relative Risk, and these also can be derived from the 2x2 contingency table we discussed in the relative risk section.

- a = Patients with the disease pattern and target disease.
- b = Patients with the disease pattern but without the target disease.
- c = Patients without the disease pattern but with the target disease.
- d = Patients without the disease pattern and target disease.

Thus,

- a + b = Total patients with disease pattern
- c + d = Total patients without disease pattern

Confidence interval:

$$CI = \left(e^{ln(RR) \pm Z \times SE}\right)$$

We used the 95% z-value to find the confidence interval, which is 1.96 from the standard normal distribution [15].

So,

$$95\%\ CI = \left(e^{ln(RR) \pm 1.96 \times SE}\right)$$

This equation gives us the CI's lower and upper bounds, which show the range where the true RR is likely to be with 95% confidence.

3.4.3. Relative Width of CI

The Relative Width (RW) of the confidence interval shows how accurate the RR estimate is. A small RW indicates that the estimated RR is more accurate and has less uncertainty around it. A wide RW represents less accuracy, leading to less reliability in RR. The reason for choosing the RW measure is that it can compare how accurate RR estimates are for different disease patterns. In our study, we calculate RW using this equation:

$$RW = \frac{upper\ limit\ of\ CI - lower\ limit\ of\ CI}{RR}$$

In 1998, O'Neill also used a similar approach in his vaccine efficiency study [16]. He calculated the relative width of their Vaccine Efficiency (VE) by dividing the CI width of VE. O'Neill then categorized the width of RW and precision of VE using threshold methods (i.e., 0.4 or 40% and 0.8 or 80%). Later in 2023, Bose and Biswas [17] followed O'Neill's lead by using the RW thresholds of 0.5 (or 50%) and 1.0 (or 100%) for their vaccine efficiency studies with different sample size groups. Accordingly, we used the RW thresholds of 0.5 and 1.0 to categorize patterns in our study.

3.5. Result Visualization

We visualized the results in tabular format on a web page. The results were stored in JSON format and shown on the web page for all unique diseases in both levels of the ICD-10 code. The webpage was built using HTML, CSS, and JavaScript, with development tools like VS Code. The table on the webpage for each disease contains these features:

- Pattern
- Patients with the disease pattern and target disease
- Percentage of patients with the disease pattern and target disease
- Total patients with disease pattern
- Percentage of total patients with disease pattern
- Relative Risk
- Confidence Interval
- Relative Width

We used conditional cell coloring based on the values in these cells to make them easier to understand.

a. Relative Risk (RR):
    - Red if the values are above 1, with darker red indicating a strong association.
    - Blue if the values are below 1, with darker blue indicating weak or no association.

b. Confidence Interval (CI):
    - Green if the lower and upper bounds are above 1.
    - Grey if the lower bound is below 1.
    - White if both bounds are below 1

c. Relative Width (RW):
    - Green if the values are between 0 and 0.5
    - Yellow for values between 0.5 and 1.0
    - Red for values above 1

4. Results and Discussions

In this study, we successfully identified that single or multiple diseases can lead to additional diseases by either a direct influence or an indirect impact. We ensured the generated patterns are predictive rather than coincidental by maintaining disease extraction that only occurred before the target disease and systematically identifying temporal relationships. We found that some patterns make meaningful clinical sense by having a higher RR value, narrow CIs, and small relative width. Some diseases produced less significant patterns due to either a low dataset for this specific disease or the disease being too vague, resulting in numerous false positive patterns with a lower relative risk and less accuracy. Mainly at the single code level, the diseases have less accurate patterns compared to the block level. This happens because single-level code is more disease-specific and can have a lower frequency due to its niche nature. Some correlation of disease and pattern often has a third factor, which is sometimes invisible in the normal sense. For example, heart failure can cause a lack of appetite, which can lead to anaemia. In our results, we can see that heart failure leads to anaemia, but the third factor, lack of appetite, is not visible. We also see that some disease and pattern relations are reversed due to the timing of entry or diagnosis. The complete result can be found here: **https://cs.uef.fi/ml/impro/disease-pattern/**

Below, we are discussing some critical disease progression for single and block levels.

4.1. Single code level

4.1.1. H36 Retinal disorders in diseases classified elsewhere

The ICD-10 code H36 signifies "Retinal disorders in diseases classified elsewhere," indicating a problem with the retina that results from another systemic disease, such as diabetic retinopathy (H36.0) or sickle-cell retinopathy. This code is a general category that requires a more specific code for the underlying condition to be used for proper diagnosis and reimbursement purposes. In our study, H36 is found in 21123 people; some important patterns for H36 are:

Pattern	H36 with Pattern	Total People with Pattern	Relative Risk	Confidence Interval	Relative Width
E10	E10, H36 = 6838	37611	48.38	(47.09, 49.70)	0.054
N08	N08, H36 = 1094	9698	21.56	(20.36, 22.84)	0.115
E11	E11, H36 = 8072	319500	6.81	(6.63, 7.00)	0.055
E10, I10	E10, I10, H36 = 1416	9664	28.47	(27.08, 29.93)	0.100
E10, K02	E10, K02, H36 = 1414	12114	22.66	(21.54, 23.85)	0.102
E11, I25	E11, I25, H36 = 1327	59662	4.25	(4.02, 4.49)	0.110
E11, I10	E11, I10, H36 = 3621	196618	3.83	(3.70, 3.97)	0.071
E11, E78	E11, E78, H36 = 1531	93022	3.16	(2.99, 3.31)	0.103
E11, K02	E11, K02, H36 = 1309	88956	2.78	(2.63, 2.94)	0.111
E11, E78, I10	E11, E78, I10, H36 = 1232	74982	3.11	(2.94, 3.29)	0.114

Table 11: Generated patterns for single level code H36

According to the analysis, Type 1 diabetes (E10) and Type 2 diabetes (E11) show strong links with H36, and the signal is especially high for E10 with an RR of 48.38. Both diabetes patterns have narrow confidence intervals and low relative width, so these estimates are stable. This observation is consistent with normal clinical practice, which identifies diabetes as the recognized cause of diabetic retinopathy (known as H36.0) [18]. We also see a strong pattern involving N08, which shows a relative risk of 21.56 when paired with H36. This N08 refers to glomerular disorders caused by other diseases instead of primary kidney problems. Since the kidney and the retina have similar small blood vessel structures, they are vulnerable to the same risks factors like diabetes, hypertension, oxidative stress, and inflammation. Factors like these

can damage both organs simultaneously. This explains why N08 often appears with H36 in our results [19]. There are also specific examples, such as C3 glomerulopathy, which is associated with drusen and other retinal changes [20].

We also see diabetes pairing with cardiovascular and metabolic conditions around retinal disease. Essential hypertension (I10), chronic ischemic heart disease (I25), and disorders of lipoprotein metabolism (E78) co-occur with H36 in several patterns. E78 is a good example of how lipids can affect the eye: lipid deposition and oxidative stress can disrupt retinal structure and function and contribute to age-related macular degeneration, diabetic retinopathy, and retinal vascular occlusive disease [21].

4.1.2. N17 Acute Kidney Failure

The ICD-10 code N17 is the general code for acute kidney failure (also known as acute kidney injury). This code serves as a parent category, with more specific sub-codes for different causes or types of acute kidney failure, such as N17.0 for acute renal failure with tubular necrosis or N17.9 for unspecified acute kidney failure. In our single-code analysis, we identified several high relative risk patterns involving single and multiple items. Some high-risk import patterns are:

Pattern	N17 with Pattern	Total People with Pattern	Relative Risk	Confidence Interval	Relative Width
N19	N19, N17 = 1038	14520	15.13	(14.24, 16.07)	0.121
N18	N18, N17 = 1910	38282	11.02	(10.53, 11.54)	0.092
A41	A41, N17 = 1525	33817	9.76	(9.27, 10.27)	0.102
E87	E87, N17 = 1628	38818	9.12	(8.67, 9.58)	0.100
A49	A49, N17 = 1869	62240	6.58	(6.27, 6.89)	0.094
D64	D64, N17 = 1456	52506	5.95	(5.64, 6.27)	0.105

I70	I70, N17 = 1348	52354	5.49	(5.20, 5.80)	0.109
N10	N10, N17 = 1814	78325	5.03	(4.80, 5.28)	0.096
I21	I21, N17 = 1378	59378	4.95	(4.68, 5.22)	0.108
M10	M10, N17 = 1216	58706	4.38	(4.13, 4.63)	0.115
E11	E11, N17 = 5248	323580	4.11	(3.98, 4.24)	0.063
E11, I50	E11, I50, N17 = 1894	38839	10.76	(10.27, 11.27)	0.093
I10, N18	I10, N18, N17 = 1162	25347	9.74	(9.19, 10.32)	0.116
I48, I50	I48, I50, N17 = 2692	73308	8.42	(8.09, 8.77)	0.080
I25, I50	I25, I50, N17 = 1865	49382	8.3	(7.92, 8.70)	0.094
I10, I50	I10, I50, N17 = 2653	74121	8.19	(7.86, 8.52)	0.081
I50, J18	I50, J18, N17 = 1377	38797	7.6	(7.21, 8.03)	0.108
E11, J18	E11, J18, N17 = 1057	35383	6.29	(5.92, 6.69)	0.122
I48, J18	I48, J18, N17 = 1130	43781	5.45	(5.13, 5.78)	0.119
E11, I25	E11, I25, N17 = 1541	60992	5.43	(5.16, 5.72)	0.103
I10, J18	I10, J18, N17 = 1651	73572	4.84	(4.60, 5.09)	0.100
I10, I48	I10, I48, N17 = 2672	134274	4.48	(4.31, 4.67)	0.081
E11, I10	E11, I10, N17 = 3479	199999	4.05	(3.90, 4.20)	0.073
E11, I10, I50	E11, I10, I50, N17 = 1315	28068	10.03	(9.50, 10.60)	0.110
E11, I48, I50	E11, I48, I50, N17 = 1088	22979	10.02	(9.44, 10.64)	0.120
I25, I48, I50	I25, I48, I50, N17 = 1085	29074	7.89	(7.43, 8.38)	0.120

I10, I25, I50	I10, I25, I50, N17 = 1102	30702	7.59	(7.15, 8.06)	0.120
I10, I48, I50	I10, I48, I50, N17 = 1570	45401	7.48	(7.11, 7.87)	0.102

Table 12: Generated patterns for single level code N17

According to the table, N18 and N19 have a high relative risk compared to N17. N18 is chronic kidney disease and can be staged further, for example, N18.1 for stage 1, while N19 is unspecified kidney failure. N17 is acute, N18 is chronic, and N19 is unspecified, but they sit on the same spectrum. CKD can progress over time, and an acute episode can occur on top of chronic disease. This is why these codes often appear together. A41, known as "Other sepsis", also stands out. It marks a severe bloodstream infection that triggers a body-wide inflammatory response and is a well-known driver of acute kidney injury [22]. The ADQI 28th consensus describes sepsis-associated AKI as common in ICU care and notes that kidney injury can develop even without long periods of low blood pressure, driven by microcirculatory changes, inflammation, and cellular stress rather than simple ischemia [22]. In practice, A41 can precipitate N17, and in a patient who already has CKD coded as N18, the presentation is acute-on-chronic kidney failure. If the documentation is incomplete, it may be captured as N19 [22].

Inflammation-related signals extend beyond sepsis. Pneumonia coded as J18 shows several high-risk combinations, including [I50, J18] with RR 7.60, [E11, J18] with RR 6.11, [I48, J18] with RR 5.45, and [I10, J18] with RR 4.84. A49, which captures bacterial infection, also links to N17 with RR 6.58.

A clear cardiorenal cluster is present. Risks rise when heart failure coded as I50 appears with diabetes or hypertension. Examples include [E11, I50] with RR 10.76, [E11, I10, I50] with RR 10.03, and [I10, I50] with RR 8.19. Pairing heart failure with atrial fibrillation is also strong, with [I48, I50] at RR 8.42. This pattern reflects low cardiac output with venous congestion, leading to renal hypoperfusion and congestion [23-25]. Diabetes and hypertension further increase AKI risk in acute heart failure cohorts [26]. Finally, electrolyte and acid-base disorders coded as E87 carry a high relative risk with N17 at RR 9.12. These imbalances can trigger acute kidney injury, for example, through dehydration or diuretics, and they can also result from it. In this setting, E87 is a useful early risk signal [27].

4.1.3. M86 Osteomyelitis

The ICD-10 code M86 stands for Osteomyelitis, which is an infection and inflammation of the bone. This code is a broad category within the Diseases of the Musculoskeletal System and Connective Tissue chapter and includes subcategories for various types of osteomyelitis, such as acute, chronic, and unspecified. We have 3733 Osteomyelitis patients in our data. Despite this low number, we still identified several high-risk patterns for this disease, characterized by a narrow confidence interval and a low relative width. A portion of the patterns is shown in the table below:

Pattern	M86 with Pattern	Total People with Pattern	Relative Risk	Confidence Interval	Relative Width
L89	L89, M86 = 291	8328	38.88	(34.57, 43.74)	0.236
L98	L98, M86 = 217	11548	20.45	(17.85, 23.43)	0.273
N08	N08, M86 = 186	10791	18.6	(16.07, 21.53)	0.293
A41	A41, M86 = 302	35323	9.48	(8.43, 10.66)	0.235
E11	E11, M86 = 1386	324619	6.39	(5.98, 6.83)	0.133
M05	M05, M86 = 197	35243	6.01	(5.21, 6.94)	0.287
A49	A49, M86 = 340	63772	5.93	(5.31, 6.63)	0.223
L02	L02, M86 = 325	74726	4.8	(4.29, 5.38)	0.228
E10, L97	E10, L97, M86 = 188	1296	157.02	(137.03, 179.93)	0.273
E11, L97	E11, L97, M86 = 442	6067	84.84	(77.07, 93.39)	0.192
A46, L97	A46, L97, M86 = 261	4198	68.66	(60.77, 77.58)	0.245
E11, S91	E11, S91, M86 = 210	3607	63.38	(55.35, 72.56)	0.271

I10, L97	I10, L97, M86 = 388	8180	54.32	(49.01, 60.20)	0.206
A46, I70	A46, I70, M86 = 238	6078	42.94	(37.75, 48.84)	0.258
I10, S91	I10, S91, M86 = 195	5262	40.15	(34.85, 46.26)	0.284
E11, I70	E11, I70, M86 = 488	19467	29.5	(26.85, 32.41)	0.189
E11, L03	E11, L03, M86 = 233	9907	25.73	(22.57, 29.33)	0.263
E11, H36	E11, H36, M86 = 301	13381	25.07	(22.31, 28.17)	0.234
A46, E11	A46, E11, M86 = 411	19538	24.18	(21.85, 26.77)	0.203
E10, E11	E10, E11, M86 = 224	10417	23.46	(20.52, 26.81)	0.268
H36, I10	H36, I10, M86 = 223	10671	22.79	(19.93, 26.06)	0.269
E10, H36	E10, H36, M86 = 205	9967	22.32	(19.42, 25.66)	0.280
I48, I70	I48, I70, M86 = 260	15555	18.4	(16.24, 20.85)	0.251
E10, I10	E10, I10, M86 = 206	12209	18.3	(15.92, 21.04)	0.280
I50, I70	I50, I70, M86 = 220	13164	18.2	(15.90, 20.83)	0.271
I10, I70	I10, I70, M86 = 473	32385	17.05	(15.49, 18.77)	0.192
I25, I70	I25, I70, M86 = 282	19608	15.92	(14.11, 17.95)	0.242
I10, L03	I10, L03, M86 = 226	17505	14.07	(12.31, 16.08)	0.268
E11, N18	E11, N18, M86 = 192	15461	13.41	(11.61, 15.49)	0.290
A46, I10	A46, I10, M86 = 410	35500	13.22	(11.94, 14.64)	0.204
A46, I48	A46, I48, M86 = 194	18319	11.43	(9.90, 13.20)	0.289
E11, I70, L97	E11, I70, L97, M86 = 204	2989	74.18	(64.71, 85.03)	0.274
E11, I10, L97	E11, I10, L97, M86 = 273	4286	70.59	(62.63, 79.55)	0.240

I10, I70, L97	I10, I70, L97, M86 = 201	3807	57.32	(49.91, 65.84)	0.278
E11, I10, I70	E11, I10, I70, M86 = 338	14192	26.83	(24.02, 29.96)	0.222
E11, H36, I10	E11, H36, I10, M86 = 190	8405	24.44	(21.15, 28.23)	0.290
E11, I25, I70	E11, I25, I70, M86 = 190	9019	22.77	(19.71, 26.31)	0.290
A46, E11, I10	A46, E11, I10, M86 = 265	13716	21.31	(18.83, 24.12)	0.248
I10, I48, I50	I10, I48, I50, M86 = 207	46427	4.8	(4.17, 5.51)	0.280

Table 13: Generated patterns for single level code M86

In M86, the patterns are dominated by diabetic foot and peripheral arterial disease. This follows the familiar pathway that starts with a lower-limb ulcer, progresses to a deep tissue infection, and then reaches the bone [28]. The signal is strongest when diabetes (E10/E11) appears together with a lower limb ulcer (L97) and atherosclerosis (I70). We see very high relative risks for [E10, L97] (RR ≈157.02) and [E11, L97] (RR ≈84.84), with a similarly strong association for [E11, S91] (RR ≈63.38). When vascular disease is added, the risk remains high in the triple patterns, for example, [E11, I70, L97] (RR ≈74.18) and [E11, I10, L97] (RR ≈70.59). Diabetes can also make infections harder to recognize and treat in time. Poor blood flow and high glucose slow healing and can mute local signs like redness or warmth. Small ulcers stay open longer, treatment starts late, and the infection has more time to reach the bone [29].

Independent of diabetes, pressure ulcers and chronic skin ulcers contribute to a substantial risk. L89 shows RR ≈38.38, and L98 shows RR ≈20.45, which matches the clinical picture where bone infection spreads directly from a nearby wound or ulcer [30-31]. Surrounding soft-tissue infection is clear in the data as well. Cellulitis (L03) rises when paired with diabetes in [E11, L03] (RR ≈25.73) and also when paired with hypertension in [I10, L03] (RR ≈14.07). Cutaneous abscess is present too in L02 with RR ≈4.80. We also see systemic infection signals

that travel with osteomyelitis. Sepsis appears as A41 with RR ≈9.48, and bacterial infection appears as A49 with RR ≈5.93. Erysipelas (A46) is especially strong when it co-occurs with an ulcer or peripheral arterial disease, as in [A46, L97] with RR ≈68.66 and [A46, I70] with RR ≈42.94, and it is elevated with diabetes as well in [A46, E11] with RR ≈24.18. Taken together, these rows read as a local infection around a chronic ulcer that grows outward and then involves the whole system. A smaller post-procedure signal also appears in T81, which is consistent with surgical-site or hardware-related osteomyelitis.

4.1.4. E85 Amyloidosis

The ICD-10 code E85 represents systemic amyloidosis, a disorder where misfolded proteins accumulate as insoluble fibrils in tissues and organs. These amyloid deposits usually cause gradual impairment to organ function that can result in a broad range of clinical issues. In our study, we found 1,408 cases of E85. Although amyloidosis is a rare disease, it was one of the largest and most consistent effects across the data set. We identified several associations with very high relative risk, and some of them are multi-organ diseases.

Pattern	E85 with Pattern	Total People with Pattern	Relative Risk	Confidence Interval	Relative Width
I68	I68, E85 = 110	218	1492.12	(1294.16, 1720.36)	0.286
H19	H19, E85 = 88	4350	58.76	(47.45, 72.76)	0.431
N08	N08, E85 = 116	10749	31.97	(26.47, 38.63)	0.380
M05	M05, E85 = 155	35229	13.35	(11.31, 15.78)	0.335
N18	N18, E85 = 105	40597	7.54	(6.18, 9.19)	0.400
I50	I50, E85 = 161	128199	3.74	(3.17, 4.40)	0.329
I48	I48, E85 = 175	243048	2.1	(1.79, 2.46)	0.318
I48, I50	I48, I50, E85 = 79	74632	3	(2.39, 3.76)	0.457

Table 14: Generated patterns for single level code E85

The strongest signal for E85 is I68, which is a classification for cerebral amyloid angiopathy. The RR of E85 compared to I68 in our cohort was extremely high, approximately 1492.12. This is much larger than the majority of associations we found within this study, and this association is partially explained by coding practices in the ICD-10 system. Specifically, for I68.0, medical professionals are instructed to "code first underlying amyloidosis (E85)" to ensure that cerebral amyloid angiopathy will be coded and systematically co-reported with amyloidosis. Additionally, there is clinical relevance to this association because cerebral amyloid angiopathy is directly caused by amyloid deposition into the walls of cerebral blood vessels, resulting in vascular wall fragility and intracerebral hemorrhage [32–33]. Both factors contribute to the strength of this association.

A second key finding of our analysis is the association with H19, which represents ocular complications of systemic diseases. Based on our data, we estimate that E85 and H19 co-occur at a relative risk of about 57.58 (95% CI 46.50-71.30). Ocular involvement is a manifestation of amyloidosis; it is represented by the presence of amyloid deposits in various parts of the eye (conjunctiva, cornea, vitreous, and trabecular meshwork). Symptoms of ocular amyloidosis may include keratopathy, floaters, or glaucoma. Approximately 8% of patients with hereditary transthyretin amyloidosis have ocular manifestations [34-35]. While ocular involvement is not a common feature of amyloidosis, its specificity accounts for the large relative risk found.

Kidney disease is another key aspect of amyloidosis in our findings. We determined that N08 (glomerular disorders in diseases classified elsewhere) is associated with E85 at a relative risk of about 31.61 (95% CI 26.17-38.20). Glomerular amyloidosis occurs when amyloid fibrils infiltrate glomeruli and lead to proteinuria and eventually to nephrotic syndrome. Eventually, the infiltration will cause chronic kidney disease. Both the process described above and the eventual chronic kidney disease are captured in our data, as N18 (chronic kidney disease) was also found to be significantly associated with E85 at a relative risk of 7.54 (95% CI 6.18-9.19). These results represent the traditional pathway of renal amyloidosis [36].

We also detected an association of relevance with M05 (seropositive rheumatoid arthritis) that was about 13.35 (95% CI 11.31-15.78). The patterns we describe correspond to secondary or

AA amyloidosis, which develops as a consequence of prolonged inflammatory diseases (such as rheumatoid arthritis). The inflammatory process increases levels of serum amyloid A protein, which is prone to misfolding and to forming amyloid fibrils. Although the incidence of AA amyloidosis has decreased in the modern era due to the advent of biologic therapies, rheumatoid arthritis remains one of the primary historical causes of AA amyloidosis, and the relationship we report is consistent with that history [36].

Finally, the analysis highlights the cardiac involvement that defines much of the morbidity and mortality of amyloidosis. We found an association between E85 and I50 (heart failure), with a relative risk of 3.74 (95% CI 3.17-4.40), and also with I48 (atrial fibrillation), with a relative risk of 2.1 (95% CI 1.79-2.46). When the two appear together, the combined pattern of I48 and I50 with E85 increases the risk further, to about 3.0 (95% CI 2.39-3.76). These results fit the clinical phenotype of cardiac amyloidosis, where amyloid infiltration stiffens the myocardium, causing a restrictive form of heart failure, and disrupts conduction pathways, predisposing patients to atrial arrhythmias [36].

4.1.5. D51 Vitamin B12 deficiency anemia

The D51 ICD-10 code represents Vitamin B12 deficiency anemia. This is a general category, and more specific codes within D51 can provide further detail about the cause of the deficiency, such as D51.0 for intrinsic factor deficiency or D51.3 for other dietary vitamin B12 deficiency anemia. In our study, we found 10339 D51 patients. Several patterns with high relative risk are identified for D51.

Pattern	D51 with Pattern	Total People with Pattern	Relative Risk	Confidence Interval	Relative Width
D64	D64, D51 = 785	52907	5.88	(5.47, 6.32)	0.145
K29	K29, D51 = 527	35287	5.79	(5.31, 6.32)	0.174
D50	D50, D51 = 628	51047	4.8	(4.43, 5.20)	0.161
E11	E11, D51 = 2208	323930	2.95	(2.81, 3.09)	0.094

E03	E03, D51 = 938	130963	2.82	(2.64, 3.02)	0.134
I50	I50, D51 = 786	127517	2.39	(2.23, 2.57)	0.145
I10	I10, D51 = 3660	722087	2.37	(2.27, 2.46)	0.080
I48	I48, D51 = 1391	242372	2.31	(2.18, 2.44)	0.113
I25	I25, D51 = 1161	203574	2.26	(2.13, 2.40)	0.122
E11, I10	E11, I10, D51 = 1393	200498	2.83	(2.67, 2.99)	0.112
I10, I25	I10, I25, D51 = 692	119037	2.24	(2.08, 2.42)	0.154
I10, I48	I10, I48, D51 = 773	135157	2.21	(2.06, 2.38)	0.147

Table 15: Generated patterns for single level code D51

The strongest links are with other anaemia codes and with upper gastrointestinal disease. D64 and D51 co-occur with RR 5.88 (95% CI 5.47 to 6.32). In everyday practice, medical personnel often record a general anaemia code alongside the specific cause, so this pairing is expected. K29 and D51 show RR 5.79 (95% CI 5.31 to 6.32). Gastritis and duodenitis impair B12 absorption through damage to the gastric mucosa and loss or dysfunction of intrinsic factor. D50 appears with D51 at RR 4.80 (95% CI 4.43 to 5.20). Iron deficiency can sit next to B12 deficiency in long-standing gastritis and related malabsorption. These upper gastrointestinal and iron deficiency links are well described in clinical reviews [37].

Metabolic and thyroid patterns are common around D51. Type 2 diabetes (E11) and D51 have RR 2.95 (95% CI 2.81 to 3.09). Many people with type 2 diabetes use metformin, and long-term metformin lowers B12 levels in a proportion of users. Major guidelines advise periodic B12 testing during long-term use, especially when anaemia or neuropathy is present. Hypothyroidism (E03) and D51 show RR 2.82 (95% CI 2.64 to 3.02). B12 deficiency and thyroid disease frequently co-exist in routine care and in autoimmune clusters [38-39].

Cardiovascular codes show moderate but steady elevations. Heart failure, atrial fibrillation, hypertension, and coronary disease each co-occur with D51 with RRs in the range of about two

to three. Part of this reflects older age and multimorbidity in the cohort. Anaemia can also worsen breathlessness and fatigue in heart failure. In addition, Heart failure can also reduce appetite and food intake, which can lead to nutritional anaemia over time. [40-41].

4.1.6. I95 Hypotension

The ICD-10 code I95 refers to Hypotension, or low blood pressure, which falls under the category of circulatory system disorders. This code and its sub-categories are used to document various types of low blood pressure, such as idiopathic, orthostatic, or drug-induced, and are important for clinical and billing purposes. We identified 23,672 patients with I95 and several high relative-risk patterns. They read as two big stories in the data, one neurodegenerative with autonomic features and one cardiac and multimorbidity.

Pattern	I95 with Pattern	Total People with Pattern	Relative Risk	Confidence Interval	Relative Width
G20	G20, I95 = 1277	21661	10.05	(9.51, 10.61)	0.110
G30	G30, I95 = 2645	95847	4.91	(4.72, 5.11)	0.080
F00	F00, I95 = 1601	56664	4.84	(4.61, 5.09)	0.100
I48	I48, I95 = 5423	241020	4.44	(4.30, 4.57)	0.060
I69	I69, I95 = 1288	55817	3.9	(3.69, 4.12)	0.111
I63	I63, I95 = 1581	74791	3.6	(3.42, 3.79)	0.101
I10	I10, I95 = 9650	720601	2.98	(2.90, 3.06)	0.052
K59	K59, I95 = 1340	77027	2.93	(2.77, 3.09)	0.109
N40	N40, I95 = 1835	114781	2.73	(2.60, 2.86)	0.094
N39	N39, I95 = 2131	138806	2.64	(2.52, 2.76)	0.088
N30	N30, I95 = 2180	143327	2.62	(2.50, 2.73)	0.087

N10	N10, I95 = 1198	78512	2.55	(2.41, 2.70)	0.116
S06	S06, I95 = 1230	81894	2.51	(2.37, 2.66)	0.114
H25	H25, I95 = 2058	158454	2.21	(2.11, 2.31)	0.090
J18	J18, I95 = 2391	193040	2.12	(2.03, 2.21)	0.084
S01	S01, I95 = 1299	105109	2.06	(1.95, 2.18)	0.111
I49	I49, I95 = 1456	118082	2.06	(1.96, 2.18)	0.106
I25, I48	I25, I48, I95 = 1797	61439	5.05	(4.82, 5.30)	0.095
G30, I10	G30, I10, I95 = 1296	44777	4.91	(4.64, 5.19)	0.110
I25, I50	I25, I50, I95 = 1380	49370	4.75	(4.50, 5.01)	0.107
F00, G30	F00, G30, I95 = 1360	49735	4.64	(4.40, 4.90)	0.108
I48, I50	I48, I50, I95 = 1914	73273	4.52	(4.32, 4.73)	0.092
I10, I50	I10, I50, I95 = 1793	74296	4.15	(3.96, 4.35)	0.095
I10, I25	I10, I25, I95 = 2698	118132	4.05	(3.89, 4.22)	0.079
I10, I48	I10, I48, I95 = 3018	133993	4.04	(3.89, 4.20)	0.076
E11, I48	E11, I48, I95 = 1308	58561	3.78	(3.57, 3.99)	0.110
E11, I25	E11, I25, I95 = 1323	61027	3.66	(3.47, 3.87)	0.110
I10, N39	I10, N39, I95 = 1208	56624	3.59	(3.39, 3.80)	0.115
E78, I25	E78, I25, I95 = 1196	62918	3.19	(3.01, 3.38)	0.115
I10, J18	I10, J18, I95 = 1246	73684	2.84	(2.68, 3.00)	0.113
E11, I10	E11, I10, I95 = 2718	200299	2.36	(2.26, 2.45)	0.079
I10, M54	I10, M54, I95 = 1685	125308	2.27	(2.16, 2.39)	0.099

I10, M79	I10, M79, I95 = 1361	104246	2.19	(2.07, 2.31)	0.109
E78, I10	E78, I10, I95 = 2499	202456	2.12	(2.03, 2.21)	0.082

Table 16: Generated patterns for single level code I95

The strongest correlation sits next to Parkinson's disease (G20). I95 pairs with G20 with a high relative risk of 10.05, and it also rises with dementia codes such as G30 at 4.91 and F00 at 4.84. This matches what we see clinically. Neurogenic orthostatic hypotension is common in Parkinson's disease because of loss of sympathetic noradrenergic innervation, and bowel and bladder autonomic symptoms often travel with it. Cohort work also links abnormal orthostatic blood pressure drops to later dementia risk. These points help explain why hypotension clusters with Parkinson's disease, Alzheimer-type dementia, and stroke outcomes in our table, including I63 at 3.60 and I69 at 3.90. Taken together, the neurodegenerative-autonomic cluster in our data is expected and biologically coherent [42-45].

The cardiac and rhythm side is on the second cluster. I95 rises with atrial fibrillation I48 at 4.44 and shows stronger signals when rhythm or coronary disease codes appear together, for example, I25 with I48 at 5.05 and I48 with I50 at 4.52. In practice, fast or irregular rhythms, rate-control drugs, diuretics, and guideline-directed heart failure therapy can all lower blood pressure. The I10 pairing at 2.98 shows that many hypotensive patients also carry a hypertension history, which fits an older, treated population rather than a single disease pathway.

A smaller but consistent genitourinary cluster shows up as well. I95 co-occurs with urinary tract and bladder diagnoses, such as N39 at 2.64, N30 at 2.62, and N10 at 2.55, as well as with constipation (K59) at 2.89. Autonomic dysfunction in synucleinopathies commonly involves the gut and lower urinary tract, so bowel dysmotility and urinary retention or incontinence can appear alongside orthostatic symptoms and may predispose to infections. This matches the table patterns we see here [44].

Finally, the data carry a frailty and falls signal. We see head injury S06 at 2.51 and open wounds of the head S01 at 2.06, along with pneumonia J18 at 2.12, cataract H25 at 2.21. Orthostatic

hypotension is linked to falls in older adults, so the injury patterns likely reflect downstream consequences rather than separate diseases. The overall picture is a plausible chain from low blood pressure to dizziness to falls and injury in a multimorbid population [46].

4.1.7. T36 Poisoning by systemic antibiotics

The ICD-10 code T36 captures poisoning by systemic antibiotics, including intentional overdoses, accidental ingestions, and other adverse toxicity events attributed to antibacterial agents. In our single-code analysis, we identified 16,689 people with T36. The highest-risk patterns cluster around mental and behavioural disorders and substance use, a profile that matches the epidemiology of deliberate self-poisoning seen in emergency departments. [47]

Pattern	T36 with Pattern	Total People with Pattern	Relative Risk	Confidence Interval	Relative Width
F19	F19, T36 = 1172	10618	27.23	(25.74, 28.80)	0.113
F60	F60, T36 = 1449	24810	14.62	(13.87, 15.40)	0.105
F10	F10, T36 = 3757	88651	12.29	(11.86, 12.74)	0.071
F31	F31, T36 = 1294	29134	10.99	(10.40, 11.62)	0.111
F33	F33, T36 = 2089	78057	6.89	(6.59, 7.21)	0.091
F32	F32, T36 = 3927	186786	6.02	(5.81, 6.23)	0.071
F43	F43, T36 = 1184	82516	3.48	(3.28, 3.69)	0.117
F51	F51, T36 = 945	86646	2.6	(2.43, 2.77)	0.131
F10, F41	F10, F41, T36 = 980	12870	18.54	(17.43, 19.73)	0.124
F10, F32	F10, F32, T36 = 1177	16770	17.29	(16.33, 18.31)	0.115
F10, K02	F10, K02, T36 = 1028	28911	8.65	(8.13, 9.20)	0.124
F32, F33	F32, F33, T36 = 1096	38154	7	(6.59, 7.44)	0.121

F32, F41	F32, F41, T36 = 1399	51327	6.75	(6.40, 7.13)	0.108
F32, K02	F32, K02, T36 = 1085	70852	3.7	(3.48, 3.93)	0.122
F41, K02	F41, K02, T36 = 877	59902	3.5	(3.27, 3.74)	0.135

Table 17: Generated patterns for single level code T36

The strongest patterns are with other psychoactive substance use F19 with RR 27.23, personality disorders F60 with RR 14.62, and alcohol use disorders F10 with RR 12.29. Mood-disorder codes also rise, bipolar disorder F31 with RR 10.99, recurrent depressive disorder F33 with RR 6.89, and depressive episode F32 with RR 6.02. These associations are clinically plausible. Hospital datasets repeatedly show that self-poisoning is the predominant method of hospital-presenting self-harm, with medications such as psychotropics and analgesics most frequently ingested [48-49]. Alcohol is commonly co-ingested and amplifies risk and severity in self-harm presentations, which matches the high F10 signal in our table [50]. Personality disorder is over-represented in deliberate self-poisoning cohorts, though intent may be lower than in other diagnostic groups, again consistent with strong, specific correlations without implying a single causal pathway [51]. The elevation with bipolar disorder mirrors registry and cohort work, where self-poisoning is a frequent attempt method in bipolar illness [52].

Combined rows reinforce the same story. Alcohol use disorder with anxiety [F10, F41] shows a relative risk of 18.54, and Alcohol use disorder with depression [F10, F32] shows a relative risk of 17.29. These stacked patterns fit the clinical setting where mood/anxiety symptoms and alcohol co-occur around overdose presentations [48-50]. We also see dental morbidities (for example, K02 dental caries) appearing in combinations like [F10, K02] with RR 8.65, which likely reflects social and behavioural comorbidity rather than a direct toxicological link.

T-chapter poisoning codes (T36-T50) are widely used to flag medication-related toxicity and adverse drug events in administrative data. Their presence indicates a drug-harm context at the encounter, but not intent or exact dose. That context helps explain why our T36 rows sit next to mental-health and alcohol codes, aligning our real-world signals with what prior clinical and pharmaco-epidemiologic studies report for overdose populations [47-50].

4.2. Block level

4.2.1. J60-J69 Lung diseases due to external agents

In this section, J60 means the whole block J60 to J69, named Lung diseases due to external agents. The block brings together long-standing dust diseases and acute conditions from inhaled chemical fumes and vapors, and pneumonitis due to solids and liquids. We identified 11105 people with a J60 block (J60-J69) in our dataset. At the block level, patterns can be recorded before or after each other in the same episode, so we read them as correlations anchored in known clinical pathways.

Pattern	J60 with Pattern	Total People with Pattern	Relative Risk	Confidence Interval	Relative Width
J90	J90, J60 = 646	23155	10.18	(9.41, 11.01)	0.157
J09	J09, J60 = 3148	248961	5.7	(5.47, 5.94)	0.082
G20	G20, J60 = 567	48035	4.25	(3.90, 4.62)	0.168
G30	G30, J60 = 1059	104429	3.77	(3.54, 4.01)	0.126
N10	N10, J60 = 857	90386	3.47	(3.24, 3.72)	0.139
N17	N17, J60 = 562	64418	3.12	(2.87, 3.40)	0.169
F10	F10, J60 = 935	120013	2.85	(2.66, 3.05)	0.133
I70	I70, J60 = 593	90377	2.34	(2.15, 2.54)	0.165
I60, J09	I60, J09, J60 = 710	28186	9.23	(8.57, 9.95)	0.150

F00, J09	F00, J09, J60 = 601	29234	7.46	(6.87, 8.09)	0.163
G40, J09	G40, J09, J60 = 652	37774	6.28	(5.80, 6.79)	0.157
I30, I60	I30, I60, J60 = 896	56593	5.87	(5.48, 6.28)	0.136
A30, J09	A30, J09, J60 = 740	47076	5.75	(5.34, 6.19)	0.148
J09, N30	J09, N30, J60 = 659	41888	5.72	(5.29, 6.18)	0.156
I10, I60	I10, I60, J60 = 1082	81502	4.98	(4.68, 5.30)	0.125
I30, J09	I30, J09, J60 = 1125	85004	4.98	(4.68, 5.29)	0.123
E70, J09	E70, J09, J60 = 574	43878	4.71	(4.34, 5.12)	0.167
I10, J09	I10, J09, J60 = 1161	93120	4.7	(4.42, 4.99)	0.121
J09, J40	J09, J40, J60 = 702	55047	4.64	(4.30, 5.00)	0.152
E10, J09	E10, J09, J60 = 569	46167	4.44	(4.08, 4.82)	0.168
I20, J09	I20, J09, J60 = 586	47736	4.42	(4.07, 4.81)	0.166
F00, I10	F00, I10, J60 = 595	57483	3.72	(3.43, 4.04)	0.165
F00, G30	F00, G30, J60 = 621	61362	3.65	(3.36, 3.95)	0.161
A30, I30	A30, I30, J60 = 584	59840	3.51	(3.23, 3.81)	0.166
I30, J40	I30, J40, J60 = 662	69347	3.45	(3.19, 3.73)	0.156
J09, K00	J09, K00, J60 = 846	95234	3.24	(3.02, 3.48)	0.140
I20, I30	I20, I30, J60 = 980	119668	3.01	(2.82, 3.21)	0.131
A30, I10	A30, I10, J60 = 593	74317	2.86	(2.63, 3.10)	0.165
G40, I30	G40, I30, J60 = 565	72723	2.78	(2.55, 3.02)	0.169
E10, I30	E10, I30, J60 = 738	98541	2.7	(2.51, 2.91)	0.149
I10, I30	I10, I30, J60 = 1560	228982	2.58	(2.44, 2.72)	0.107

I10, J40	I10, J40, J60 = 722	105484	2.46	(2.28, 2.65)	0.151
I10, I20	I10, I20, J60 = 906	140533	2.34	(2.18, 2.50)	0.135
E70, I30	E70, I30, J60 = 645	102504	2.25	(2.08, 2.43)	0.159

Table 18: Generated patterns for block J60-J69

The clearest pattern is with J90 pleural disease. The relative risk for J90 with J60 to J69 is 10.18 with a confidence interval of 9.41 to 11.01. This matches the asbestos literature where pleural plaques, effusions, and diffuse thickening frequently co-occur with pneumoconioses and parenchymal fibrosis, which explains a tight clinical correlation in exposed populations [53].

Influenza and pneumonia J09 also correlate strongly with this block. The relative risk is 5.7 with a confidence interval of 5.47 to 5.94. Combination patterns tell the same story with I60 and J09 at 9.23, with I30 and J09 at 4.98, with I10 and J09 at 4.7, and with J40 and J09 at 4.64. Viral infection is a recognised trigger for acute worsening in interstitial lung disease and for exacerbations in chronic obstructive pulmonary disease, which is consistent with these correlations at the block level [54, 55].

Neurological patterns fit an aspiration-prone phenotype within this block through pneumonitis due to solids and liquids. G20 shows 4.25, and G30 shows 3.77, and F00 (dementia) with J09 shows 7.46. Dysphagia after stroke and impaired airway protection in Parkinson's disease are both well linked to aspiration pneumonia, which supports what we observe in the table [56-57].

Cardiometabolic and renal codes read as background vulnerability rather than a specific exogenous cause. I70 shows 2.34, N47 shows 3.47, and N17 shows 3.12. Population data associate hypertension with a higher risk of incident pneumonia and more severe respiratory courses, which can raise co-coding during acute admissions and align with the correlations we see here [58].

Taken together, the block-level view resolves two coherent pictures. One is a pleural and parenchymal pattern that tracks with occupational dust exposure. The other is an aspiration-prone pattern in which neurological disease and seasonal viral infection appear together.

4.2.2. I60-I69 Cerebrovascular diseases

Here, I60 represents the whole block I60 to I69, named Cerebrovascular diseases. We identified 137544 people with I60 block. Important patterns are as follows:

Pattern	I60 with Pattern	Total People with Pattern	Relative Risk	Confidence Interval	Relative Width
F00	F00, I60 = 10444	103785	2.96	(2.90, 3.01)	0.038
G30	G30, I60 = 8994	96818	2.7	(2.65, 2.76)	0.041
I30	I30, I60 = 34285	444657	2.53	(2.50, 2.56)	0.024
I70	I70, I60 = 7327	83733	2.52	(2.47, 2.58)	0.045
I10	I10, I60 = 48922	702033	2.47	(2.44, 2.49)	0.021
G40	G40, I60 = 21269	355382	1.79	(1.77, 1.82)	0.028
I20, I30	I20, I30, I60 = 10204	109410	2.73	(2.68, 2.78)	0.039
G40, I10	G40, I10, I60 = 9567	106291	2.62	(2.57, 2.68)	0.040
I10, I30	I10, I30, I60 = 18020	209097	2.62	(2.58, 2.66)	0.030
E10, I30	E10, I30, I60 = 8150	90855	2.6	(2.54, 2.65)	0.043
I10, I20	I10, I20, I60 = 11078	128911	2.52	(2.47, 2.57)	0.037
E70, I30	E70, I30, I60 = 6939	90063	2.21	(2.16, 2.26)	0.046
E10, I10	E10, I10, I60 = 14183	200671	2.08	(2.05, 2.12)	0.034

E70, I10	E70, I10, I60 = 14529	212142	2.02	(1.99, 2.05)	0.033

Table 19: Generated patterns for block I60-I69

The vascular background is strong and consistent in the table. Hypertensive disease coded within I10 shows a relative risk of 2.47 with a confidence interval of 2.44 to 2.49. Other forms of heart disease within I30 show 2.53 with 2.50 to 2.56. Diseases of arteries within I70 show 2.52 with 2.47 to 2.58. These values match large multi-country and meta-analytic evidence where raised blood pressure and common cardiac causes account for a major share of stroke risk across populations, and where atrial fibrillation within the I30 block is a specific and potent mechanism for embolic stroke [45-48].

Neurodegenerative patterns are also visible. Dementia within F00 sits at 2.96, with 2.90 to 3.01. Degenerative disorders within G30 sit at 2.70. This reads as shared vascular pathology and also as cognitive sequelae after cerebrovascular events. Systematic reviews describe substantial burdens of pre-stroke and post-stroke dementia, which fits the strength of these correlations in routine data [63]. Episodic neurological disorders are part of the profile. G40 shows 1.79 with 1.77 to 1.82. Post-stroke seizures and later epilepsy are well described in cohort and prediction studies, which supports this signal even though the block mixes ischaemic and haemorrhagic subtypes [64].

The combination rows keep the same story. I20 with I30 shows 2.48 with 2.43 to 2.53. I10 with I30 shows 2.37 with 2.33 to 2.41. E10 with I10 shows 1.96 with 1.93 to 1.99. These stacked patterns are expected where vascular risk clusters in the same people [59-60].

4.2.3. N17-N19 Renal failure

Here, N17 represents the whole block N17 to N19, named Renal failure. We identified 64807 people with the N17 code at the block level.

Pattern	N17 with Pattern	Total People with Pattern	Relative Risk	Confidence Interval	Relative Width
N00	N00, N17 = 3324	12241	16.9	(16.40, 17.41)	0.060
D60	D60, N17 = 4250	52995	5.01	(4.87, 5.17)	0.060
I30	I30, N17 = 24407	456823	4.47	(4.40, 4.54)	0.031
I70	I70, N17 = 5334	86462	3.89	(3.79, 4.00)	0.054
N10	N10, N17 = 5161	84976	3.82	(3.72, 3.93)	0.055
I10	I10, N17 = 30024	721447	3.73	(3.67, 3.79)	0.030
F00	F00, N17 = 4610	111345	2.56	(2.49, 2.64)	0.059
I60	I60, N17 = 5434	132883	2.55	(2.48, 2.62)	0.055
G30	G30, N17 = 4018	101907	2.42	(2.35, 2.50)	0.063
H30	H30, N17 = 4249	112881	2.32	(2.25, 2.39)	0.061
E10, I30	E10, I30, N17 = 8598	92826	6.17	(6.04, 6.31)	0.044
I30, M05	I30, M05, N17 = 3502	38041	5.71	(5.52, 5.90)	0.065
I20, I30	I20, I30, N17 = 9315	113251	5.52	(5.41, 5.64)	0.042
A30, I30	A30, I30, N17 = 4551	53371	5.36	(5.20, 5.51)	0.058
I10, I70	I10, I70, N17 = 3442	46147	4.61	(4.46, 4.76)	0.066
I10, I30	I10, I30, N17 = 13971	218680	4.55	(4.47, 4.63)	0.036
I30, J09	I30, J09, N17 = 5504	78422	4.45	(4.33, 4.57)	0.053
I10, M05	I10, M05, N17 = 4072	58810	4.31	(4.18, 4.44)	0.061
A30, I10	A30, I10, N17 = 4587	67130	4.28	(4.16, 4.41)	0.058
I10, I20	I10, I20, N17 = 8481	133942	4.16	(4.07, 4.26)	0.044

E10, I10	E10, I10, N17 = 12154	204886	4.09	(4.02, 4.17)	0.038
I30, J40	I30, J40, N17 = 4317	66498	4.05	(3.93, 4.17)	0.060
I30, N30	I30, N30, N17 = 3990	66588	3.72	(3.60, 3.83)	0.062
E70, I30	E70, I30, N17 = 5624	96278	3.69	(3.60, 3.79)	0.053
I10, J09	I10, J09, N17 = 4853	86107	3.53	(3.43, 3.63)	0.057
E70, I20	E70, I20, N17 = 3893	77629	3.1	(3.00, 3.20)	0.063
E10, E70	E10, E70, N17 = 5188	107557	3.02	(2.94, 3.10)	0.055
I10, I60	I10, I60, N17 = 3718	77961	2.94	(2.84, 3.03)	0.065
I10, N30	I10, N30, N17 = 4418	94861	2.89	(2.80, 2.97)	0.060
I10, J40	I10, J40, N17 = 4633	102238	2.81	(2.73, 2.90)	0.059
I30, M15	I30, M15, N17 = 3701	83005	2.74	(2.65, 2.83)	0.065
E70, I10	E70, I10, N17 = 8748	224479	2.51	(2.46, 2.57)	0.044
I10, K55	I10, K55, N17 = 3513	87888	2.45	(2.37, 2.53)	0.067
H25, I10	H25, I10, N17 = 3296	85829	2.34	(2.26, 2.42)	0.069
I30, M70	I30, M70, N17 = 3861	105292	2.25	(2.18, 2.32)	0.064
E10, I20, I30	E10, I20, I30, N17 = 3948	34724	7.11	(6.89, 7.33)	0.061
E10, I10, I30	E10, I10, I30, N17 = 5975	66419	5.77	(5.62, 5.92)	0.051
I10, I20, I30	I10, I20, I30, N17 = 5683	69251	5.23	(5.10, 5.37)	0.052
E10, I10, I20	E10, I10, I20, N17 = 4048	48341	5.22	(5.07, 5.39)	0.061
I10, I30, J09	I10, I30, J09, N17 = 3360	44331	4.68	(4.53, 4.84)	0.067

E70, I10, I30	E70, I10, I30, N17 = 4267	70792	3.75	(3.64, 3.87)	0.060
E10, E70, I10	E10, E70, I10, N17 = 4266	84499	3.13	(3.04, 3.23)	0.061

Table 20: Generated patterns for block N17-N19

The kidney-intrinsic patterns are the most pronounced, and they mirror what we reported in the single-code N17 analysis. Glomerular disease N00 shows a very high relative risk of 16.09 with a confidence interval of 16.40 to 17.41, which is compatible with abrupt loss of filtration from immune glomerulonephritis in clinical cohorts [65]. Tubulo-interstitial processes N10 rises to 3.82, with 3.72 to 3.93, which is consistent with drug-related and infectious interstitial nephritis and acute pyelonephritis as common contexts for acute kidney injury [65, 66]. In the single-code table, N18 and N19 were also strongly linked to N17, with relative risks around 10 for N18 and 14 for N19, which reflects the clinical spectrum where acute injury appears on top of chronic disease or is captured as unspecified kidney failure at admission. Infection was prominent at the single-code level through A41 sepsis and E87 electrolyte and acid–base disturbance. The block view preserves the same story through the N10 signal and through raised risks in rows that include acute respiratory infection, which tracks with hospital data on infection and sepsis as leading settings for acute kidney injury [67, 68, 66].

A broad cardiovascular cluster sits alongside these kidney drivers, and it is coherent with a cardiorenal framework. Other heart disease I30 shows 4.57 with 4.40 to 4.54. Hypertensive disease I10 shows 3.73 with 3.67 to 3.79. Diseases of arteries I70 show 3.89 with 3.79 to 4.00. Combination rows keep the same picture. Diabetes with heart disease E10 with I30 is 6.17 with 6.04 to 6.31. I10 with I30 is 4.55 with 4.47 to 4.63. These stacked patterns match the description of type 1 and type 2 cardiorenal interactions, where acute or chronic cardiac dysfunction and vascular disease heighten the risk of acute kidney injury through haemodynamic shifts, venous congestion, and neurohormonal activation [69].

Neurological and vascular codes appear as well, and they fit known overlaps from stroke cohorts. Cerebrovascular disease I60 sits at 2.55 with 2.48 to 2.62, and dementia F00 sits at 2.56 with 2.49 to 2.64. Meta-analytic work shows that acute kidney injury is a frequent

complication after stroke and is tied to worse outcomes, which supports these correlations in routine data [70].

4.2.4. T36-T50 Poisoning by drugs, medicaments, and biological substances

At the block level, T36 stands for the whole block T36-T50, which covers poisoning by prescribed and over-the-counter drugs, medicaments, and biological substances. We identified 18,679 people with a T36–T50 code at the block level. These codes flag a drug-harm context at the encounter and do not, on their own, specify intent or exact dose. In our data, the highest-risk patterns cluster around mental and behavioral disorders and alcohol, which mirrors what emergency and registry studies report for hospital-presenting self-poisoning.

Pattern	T36 with Pattern	Total People with Pattern	Relative Risk	Confidence Interval	Relative Width
F30	F30, T36 = 6406	257992	7.24	(7.03, 7.46)	0.060
F20	F20, T36 = 1700	60459	6.26	(5.96, 6.57)	0.098
F50	F50, T36 = 1354	99963	2.92	(2.77, 3.09)	0.110
F10, F30	F10, F30, T36 = 2460	31647	18.25	(17.51, 19.01)	0.082
F30, F60	F30, F60, T36 = 1368	25198	11.96	(11.33, 12.62)	0.107
F40, F60	F40, F60, T36 = 1004	19866	10.92	(10.26, 11.62)	0.124
F10, K00	F10, K00, T36 = 2368	56575	9.7	(9.30, 10.12)	0.085
F10, M50	F10, M50, T36 = 1117	25412	9.54	(9.00, 10.13)	0.119
F10, S00	F10, S00, T36 = 1215	29714	8.92	(8.42, 9.44)	0.114
F30, F40	F30, F40, T36 = 2824	101299	6.57	(6.32, 6.84)	0.079

F30, M50	F30, M50, T36 = 1286	64255	4.34	(4.11, 4.59)	0.112
F30, K00	F30, K00, T36 = 2602	141633	4.22	(4.05, 4.40)	0.082
F30, J00	F30, J00, T36 = 1050	61966	3.63	(3.41, 3.86)	0.124
F30, M70	F30, M70, T36 = 1027	62613	3.51	(3.30, 3.73)	0.125
F40, M50	F40, M50, T36 = 1054	69076	3.26	(3.07, 3.47)	0.124
F40, K00	F40, K00, T36 = 2043	168610	2.67	(2.55, 2.80)	0.092
F40, J00	F40, J00, T36 = 935	75429	2.63	(2.46, 2.81)	0.131
F10, F30, F40	F10, F30, F40, T36 = 1216	14715	18.09	(17.11, 19.13)	0.112
F10, F30, K00	F10, F30, K00, T36 = 1115	18283	13.26	(12.51, 14.07)	0.118
F10, F40, K00	F10, F40, K00, T36 = 968	16058	13.01	(12.22, 13.85)	0.126
F30, F40, K00	F30, F40, K00, T36 = 1275	60751	4.56	(4.31, 4.82)	0.113

Table 21: Generated patterns for block T36-T50

The strongest single-block correlations sit with mood disorders and psychotic disorders. F30 Mood [affective] disorders show RR of 7.24 with 95% CI 7.03 to 7.46, and F20 Schizophrenia, schizotypal and delusional disorders show RR of 6.26 with 95% CI 5.96 to 6.57. F50 Eating disorders are also elevated with RR 2.92 and 95% CI 2.77 to 3.09. These signals track published cohorts where poisoning is a frequent method around self-harm presentations in mood and psychotic illness, and where eating-disorder populations carry high rates of self-injury and suicide attempts that often involve medication ingestion [71, 72-74].

The combination of rows sharpens the same story. Pairing alcohol with mood codes gives a very high risk. F10 Alcohol use disorders with F30 Mood disorders together show RR 18.25 with 95% CI 17.51 to 19.01. Stacked mood and personality patterns are also strong, for example, F30 and F60 Personality disorders with RR of 11.96 and 95% CI 11.33 to 12.62, and

F40 Anxiety disorders and F60 show RR of 10.92 and 95% CI 10.26 to 11.62. These rows read like typical self-poisoning encounters in hospital data, where mood or anxiety symptoms, personality pathology, and alcohol co-occur, and where co-ingested alcohol is common at the bedside [47-48, 71].

Beyond the core psychiatric patterns, several context codes rise when paired with mood or alcohol. F10 and S00 (Injuries to the head) show RR of 8.92 with 95% CI 8.42 to 9.44, and F10 with M50 (Other dorsopathies) shows RR of 8.93 with 95% CI 9.00 to 10.13. Dental and oral-health rows behave similarly, for example, F10 with K00 (Diseases of the oral cavity, salivary glands, and jaws) shows RR of 9.70 and 95% CI 9.30 to 10.12, and F30 and K00 show RR of 4.22. We interpret these combinations as social and behavioural comorbidity signals linked to overdose presentations rather than direct toxicological pathways. The block-level perspective is helpful here, since T36–T50 codes are widely used in administrative data to identify drug poisonings and adverse drug events, so these patterns are expected to cluster with mental-health and alcohol codes in routine care [47].

These results line up with our single-code T36 analysis, which showed the same dominant themes around alcohol, mood and anxiety disorders, and personality pathology. The block view adds depth by confirming these psychiatric codes remain as key high-risk correlates when all drug-poisoning codes are pooled, and by revealing consistent stacked patterns with alcohol and injury that fit the clinical setting of self-poisoning in emergency care [48, 71].

4.3. Understudied Patterns

We tried to find some understudied and unusual directional patterns that may need further study. Some patterns do not reflect temporal casuality and many of them seem to be symptom-disease connections. But we found some interesting patterns also; for example, migraine (G43) appeared with uterine or pelvic inflammatory diseases (N71, N73). This could be related to hormone and inflammation effects, and migraine is more likely a consequence than a cause in these cases. The table below shows this type of pattern for single-level codes.

Pattern	Target	Target with Pattern	Total People with Pattern	Relative Risk
G43	N71	G43, N71 = 608	76803	3.04

K59	C19	K59, C19 = 80	78017	2.99
N92	N70	N92, N70 = 192	55608	5.35
N40	G23	N40, G23 = 119	115796	3.27
G43	N73	G43, N73 = 252	77075	2.75
N83	N70	N83, N70 = 503	30930	29.04
M17	M18	M17, M18 = 999	207176	2.12
N92	L68	N92, L68 = 128	55645	6.38

Table 22: Understudied single-code level patterns

5. Conclusion

This study has shown that by integrating pattern discovery with statistical validation, it is possible to identify clinically relevant temporal disease associations in large electronic health record databases. The ability to analyze patterns at two levels, the single code and block levels, will help researchers not only to find the sequence of diseases but also identify trends associated with blocks of diseases. The methodology developed for the study calculated ranked pattern list, which can help medical personnel to find out the clinical relevance of the patterns and statistical reliability. The result also can support understanding the order of occurrence of the diseases and identify those patients who are at high risk. These findings provide evidence that this method could be applied to generate hypothesis and support decision-making in the area of disease prediction.

6. References

[1] D. Yach, C. Hawkes, C. L. Gould, and K. J. Hofman, “The global burden of chronic diseases,” JAMA, vol. 291, no. 21, p. 2616, Jun. 2004, doi: 10.1001/jama.291.21.2616.

[2] World Health Organization: WHO, “Noncommunicable diseases,” Dec. 23, 2024. https://www.who.int/news-room/fact-sheets/detail/noncommunicable-diseases (accessed Mar. 08, 2025).

[3] World Health Organization: WHO, “Preventing noncommunicable diseases,” Dec. 09, 2024. https://www.who.int/activities/preventing-noncommunicable-diseases (accessed Mar. 08, 2025).

[4] S. H. Woolf and H. Schoomaker, “Life expectancy and mortality rates in the United States, 1959-2017,” JAMA, vol. 322, no. 20, p. 1996, Nov. 2019, doi: 10.1001/jama.2019.16932.

[5] Uddin S, Wang S, Khan A, Lu H. Comorbidity progression patterns of major chronic diseases: The impact of age, gender and time-window. Chronic Illness [Internet]. 2022 Mar 21;19(2):304–13. Available from: https://doi.org/10.1177/17423953221087647

[6] Su X, Ren Y, Li M, Zhao X, Kong L, Kang J. Prevalence of comorbidities in asthma and nonasthma patients. Medicine [Internet]. 2016 May 1;95(22):e3459. Available from: https://doi.org/10.1097/md.0000000000003459

[7] Rocca E, Anjum RL. Complexity, Reductionism and the Biomedical Model. In: Rethinking Causality, Complexity and Evidence for the Unique Patient [Internet]. 2020. p. 75–94. Available from: https://doi.org/10.1007/978-3-030-41239-5_5

[8] Elewonibi B, Nkwonta C. The association of chronic diseases and mammography among Medicare beneficiaries living in Appalachia. Women S Health [Internet]. 2020 Jan 1;16. Available from: https://doi.org/10.1177/1745506520933020

[9]. Han J, Pei J, Yin Y. Mining frequent patterns without candidate generation. ACM SIGMOD Record [Internet]. 2000 May 16;29(2):1–12. Available from: https://doi.org/10.1145/335191.335372

[10] J. Zhang and K. F. Yu, "What's the relative risk?," JAMA, vol. 280, no. 19, p. 1690, Nov. 1998, doi: 10.1001/jama.280.19.1690.

[11] K. J. Rothman, S. Greenland, and T. L. Lash, Modern Epidemiology, 3rd ed., Philadelphia, PA: Lippincott Williams & Wilkins, 2008, p. 53.

[12] D. Katz, J. Baptista, S. P. Azen, and M. C. Pike, "Obtaining confidence intervals for the risk ratio in cohort studies," Biometrics, vol. 34, no. 3, p. 469, Sep. 1978, doi: 10.2307/2530610.

[13] A. Agresti and B. A. Coull, "Approximate is Better than 'Exact' for Interval Estimation of Binomial Proportions," The American Statistician, vol. 52, no. 2, pp. 119–126, May 1998, doi: 10.1080/00031305.1998.10480550.

[14] A. Wald, "Tests of Statistical Hypotheses Concerning Several Parameters When the Number of Observations is Large," Transactions of the American Mathematical Society, vol. 54, no. 3, p. 426, Nov. 1943, doi: 10.2307/1990256.

[15] N. W. P. E, "Statistical Methods for Research Workers. By R. A. Fisher. [Pp. 239 + ix + VI Tables. Edinburgh and London: Oliver & Boyd. 1925. Price 15s.] - The Fundamentals of Statistics. By L. L. Thurstone. [Pp. 237 + xvi. New York: The Macmillan Company. 1925. Price 8s. 6d.]," Journal of the Institute of Actuaries, vol. 56, no. 3, pp. 326–327, Nov. 1925, doi: 10.1017/s0020268100030833.

[16] O'neill RT. On sample sizes to estimate the protective efficacy of a vaccine. Statistics in Medicine [Internet]. 1988 Dec 1;7(12):1279–88. Available from: https://doi.org/10.1002/sim.4780071208

[17] Bose M, Biswas A. Sample sizes required to estimate the protective efficacy of a vaccine when there is an unequal allocation of individuals across the vaccine and placebo groups. Statistical Methods in Medical Research [Internet]. 2023 Aug 30;32(10):1859–79. Available from: https://doi.org/10.1177/09622802231176807

[18] Chung YR, Ha KH, Lee K, Kim DJ. Diabetic retinopathy and related clinical practice for people with diabetes in Korea: a 10-year trend analysis. Diabetes Metab J. 2020;44(6):928-932. doi:10.4093/dmj.2020.0096.

[19] Wong CW, Wong TY, Cheng CY, Sabanayagam C. Kidney and eye diseases: common risk factors, etiological mechanisms, and pathways. Kidney International [Internet]. 2013 Dec 11;85(6):1290–302. Available from: https://doi.org/10.1038/ki.2013.491

[20] Savige J, Amos L, Ierino F, Mack HG, Symons RCA, Hughes P, et al. Retinal disease in the C3 glomerulopathies and the risk of impaired vision. Ophthalmic Genetics [Internet]. 2016 Feb 25;37(4):369–76. Available from: https://doi.org/10.3109/13816810.2015.1101777

[21] Song W, Chen W, Chi J, Liu X, Zhu W. The role of lipid metabolism disorder in the progression and treatment of ocular vascular diseases. Survey of Ophthalmology [Internet]. 2025 Sep 1; Available from: https://doi.org/10.1016/j.survophthal.2025.09.002

[22] Zarbock A, Nadim MK, Pickkers P, Gomez H, Bell S, Joannidis M, et al. Sepsis-associated acute kidney injury: consensus report of the 28th Acute Disease Quality Initiative workgroup. Nat Rev Nephrol. 2023;19(7):401-417.

[23] Mullens W, Abrahams Z, Francis GS, Sokos G, Taylor DO, Starling RC, et al. Importance of venous congestion for worsening of renal function in advanced decompensated heart failure. J Am Coll Cardiol. 2009;53(7):589–596. doi:10.1016/j.jacc.2008.05.068.

[24] Damman K, Navis G, Smilde TD, Voors AA, van der Bij W, van Veldhuisen DJ, et al. Decreased cardiac output, venous congestion and the association with renal impairment in patients with cardiac dysfunction. Eur J Heart Fail. 2007;9(9):872–878. doi:10.1016/j.ejheart.2007.05.010.

[25] Rangaswami J, Bhalla V, Blair JEA, Chang TI, Costa S, Lentine KL, et al. Cardiorenal syndrome: classification, pathophysiology, diagnosis, and treatment strategies: A Scientific Statement from the American Heart Association. Circulation. 2019;139(16):e840–e878. doi:10.1161/CIR.0000000000000664.

[26] Chen JJ, Lee TH, Kuo G, Yen CL, Chen SW, Chu PH, et al. Acute kidney disease after acute decompensated heart failure. Kidney Int Rep. 2022;7(3):526–536. doi:10.1016/j.ekir.2021.12.033.

[27] Chen X, Xu J, Li Y, Xu X, Shen B, Zou Z, et al. Risk scoring systems including electrolyte disorders for predicting the incidence of acute kidney injury in hospitalized patients. Clinical Epidemiology [Internet]. 2021 May 1;Volume 13:383–96. Available from: https://doi.org/10.2147/clep.s311364

[28] Giurato L, Meloni M, Izzo V, et al. Osteomyelitis in diabetic foot: a comprehensive overview. Diabetes Metab Res Rev. 2017;33(Suppl 1):e2924. (Open-access summary on DFO pathophysiology/contiguous spread.)

[29] Senneville É, Albalawi Z, van Asten SAV, Abbas ZG, Allison G, Aragón-Sánchez J, et al. IWGDF/IDSA guidelines on the diagnosis and treatment of foot infection in persons with diabetes. Clin Infect Dis. 2023. doi:10.1093/cid/ciad527.

[30] Wong D, Holtom P, Spellberg B. Osteomyelitis complicating sacral pressure ulcers: systematic review. Clin Infect Dis. 2018;68(2):338–345.

[31] Chicco M, et al. Diagnosing pelvic osteomyelitis in pressure ulcers: systematic review. J Wound Care. 2020;29(Sup9b):S4–S11.

[32] Biffi A, Greenberg SM. Cerebral amyloid angiopathy: a systematic review. J Clin Neurol. 2011;7(1):1–9.

[33] Charidimou A, Boulouis G, Frosch MP, et al. Emerging concepts in sporadic cerebral amyloid angiopathy. Brain. 2017;140(7):1829–50.

[34] Mathew R. Ocular amyloidosis. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2024.

[35] Gondim FA, Gonçalves JA, Oliveira AS, et al. Ophthalmological manifestations of hereditary transthyretin amyloidosis. Arq Bras Oftalmol. 2022;85(6):1–8.

[36] Merlini G, Bellotti V. Molecular mechanisms of amyloidosis. N Engl J Med. 2003;349(6):583–96.

[37] Kulnigg-Dabsch S. Autoimmune gastritis. Wien Med Wochenschr. 2016;166(13-14):424-430.

[38] de Jager J, Kooy A, Lehert P, et al. Long term treatment with metformin in patients with type 2 diabetes and risk of vitamin B-12 deficiency, randomised placebo controlled trial. BMJ. 2010;340:c2181.

[39] American Diabetes Association Professional Practice Committee. Standards of Care in Diabetes 2024. Diabetes Care. 2024;47(Suppl 1):S158-S201 and related sections.

[40] Beverborg NG, van der Wal HH, Klip IT, van der Meer P. Anemia in heart failure, still relevant. JACC Heart Fail. 2018;6(3):201-208.

[41] Kida K, et al. Nutritional management of heart failure. J Cardiol. 2023;81(6):506-512. Notes loss of appetite and malnutrition in heart failure.

[42] Pfeiffer RF. Autonomic dysfunction in Parkinson's disease. Neurotherapeutics. 2020;17(4):1464-1479. doi:10.1007/s13311-020-00897-4.

[43] Cutsforth-Gregory JK, Low PA. Neurogenic orthostatic hypotension in Parkinson disease: a primer. Neurol Ther. 2019;8(2):307-324. doi:10.1007/s40120-019-00152-9.

[44] Mendoza-Velásquez JJ, Flores-Vázquez JF, Barrón-Velázquez E, Sosa-Ortiz AL, Illigens BMW, Siepmann T. Autonomic dysfunction in α-synucleinopathies. Front Neurol. 2019;10:363. doi:10.3389/fneur.2019.00363.

[45] Ma Y, Zhang Y, Coresh J, et al. Orthostatic blood pressure change, dizziness, and risk of dementia in the ARIC study. Hypertension. 2024;81(1):96-106. doi:10.1161/HYPERTENSIONAHA.123.21438.

[46] Mol A, Bui Hoang PTS, Sharmin S, et al. Orthostatic hypotension and falls in older adults: a systematic review and meta-analysis. J Am Med Dir Assoc. 2019;20(5):589-597.e5. doi:10.1016/j.jamda.2018.11.003.

[47] Hohl CM, Karpov A, Reddekopp L, Doyle-Waters M, Stausberg J. ICD-10 codes used to identify adverse drug events in administrative data: a systematic review. J Am Med Inform Assoc. 2014;21(3):547-57. doi:10.1136/amiajnl-2013-002116.

[48] Hendrix L, Verelst S, Desruelles D, Gillet JB. Deliberate self-poisoning patients presenting to the emergency department: Who they are and what they ingest. West J Emerg Med. 2013;14(4):337-44. doi:10.5811/westjem.2012.12.15114.

[49] Geulayov G, Kapur N, Turnbull P, Clements C, Waters K, Ness J, et al. Epidemiology and trends in non-fatal self-harm in England: evidence from the Multicentre Study of Self-harm in England, 2000–2012. BMJ Open. 2016;6:e010538.

[50] Miller M, Borges G, Orozco R, Mukamal K, Rascati K, Haro JM, et al. Exposure to alcohol, drugs and nicotine and the risk of subsequent suicide attempts in the general population. Suicide Life Threat Behav. 2010;40(5):490-9.

[51] Grimholt TK, Jacobsen D, Haavet OR, Ekeberg Ø. Lower suicide intention in patients with personality disorders admitted for deliberate self-poisoning than in patients with other diagnoses. Ann Gen Psychiatry. 2017;16:21.

[52] Schaffer A, Isometsä ET, Tondo L, H Moreno D, Turecki G, Reis C, et al. Epidemiology, neurobiology and pharmacological interventions related to suicide deaths and suicide attempts in bipolar disorder: Part I. Front Psychiatry. 2017;8:33.

[53] American Thoracic Society. Diagnosis and initial management of nonmalignant diseases related to asbestos. Am J Respir Crit Care Med. 2004;170(6):691-715. doi:10.1164/rccm.200310-1436ST.

[54] Wootton SC, Kim DS, Kondoh Y, et al. Viral infection in acute exacerbation of idiopathic pulmonary fibrosis. Am J Respir Crit Care Med. 2011;183(12):1698-1702.

doi:10.1164/rccm.201010-1752OC.

[55] Wedzicha JA, Seemungal TAR. COPD exacerbations defining their cause and prevention. Lancet. 2007;370(9589):786-796. doi:10.1016/S0140-6736(07)61382-8.

[56] Chang MC, Park D. The relationship between dysphagia and pneumonia in acute stroke patients a systematic review and meta-analysis. Medicina. 2022;58(3):311. doi:10.3390/medicina58030311.

[57] Won JH, Byun SJ, Kim HJ, et al. Risk and mortality of aspiration pneumonia in Parkinson's disease a nationwide database study. Sci Rep. 2021;11:6592. doi:10.1038/s41598-021-86011-w.

[58] Zekavat SM, Honigberg MC, Pirruccello JP, et al. Elevated blood pressure increases pneumonia risk epidemiological association and Mendelian randomisation in the UK Biobank. Am J Hypertens. 2021;34(3):307-315. doi:10.1093/ajh/hpaa193.

[59] O'Donnell MJ, Chin SL, Rangarajan S, et al. Global and regional effects of potentially modifiable risk factors for stroke. Lancet. 2016;388(10046):761-775. doi:10.1016/S0140-6736(16)30506-2.

[60] Lewington S, Clarke R, Qizilbash N, Peto R, Collins R; Prospective Studies Collaboration. Age specific relevance of usual blood pressure to vascular mortality. Lancet. 2002;360(9349):1903-1913. doi:10.1016/S0140-6736(02)11911-8.

[61] Wolf PA, Abbott RD, Kannel WB. Atrial fibrillation as an independent risk factor for stroke. Stroke. 1991;22(8):983-988. doi:10.1161/01.STR.22.8.983.

[62] Smeeth L, Thomas SL, Hall AJ, Hubbard R, Farrington P, Vallance P. Risk of myocardial infarction and stroke after acute infection or vaccination. N Engl J Med. 2004;351(25):2611-2618. doi:10.1056/NEJMoa041747.

[63] Pendlebury ST, Rothwell PM. Prevalence incidence and factors associated with pre stroke and post stroke dementia. Lancet Neurol. 2009;8(11):1006-1018. doi:10.1016/S1474-4422(09)70236-4.

[64] Galovic M, Ferreira-Atuesta C, Abraira L, et al. Prediction of late seizures after ischaemic stroke with a simple risk score. Lancet Neurol. 2018;17(11):988-997. doi:10.1016/S1474-4422(17)30404-0.

[65] KDIGO Clinical Practice Guideline for Acute Kidney Injury. Kidney Int Suppl. 2012;2(1):1-138.

[66] Johnson JR, Russo TA. Acute pyelonephritis in adults. N Engl J Med. 2018;378(1):48-59.

[67] Hoste EAJ, Bagshaw SM, Bellomo R, et al. Epidemiology of acute kidney injury in critically ill patients the multinational AKI-EPI study. Intensive Care Med. 2015;41(8):1411-1423.

[68] ADQI 28 Workgroup. Sepsis-associated acute kidney injury consensus report. Nat Rev Nephrol. 2023;19(7):471-489.

[69] Ronco C, Haapio M, House AA, Anavekar N, Bellomo R. Cardiorenal syndrome. J Am Coll Cardiol. 2008;52(19):1527-1539.

[70] Arnold J, Ng KP, Zainuldin R, et al. Incidence and impact on outcomes of acute kidney injury after a stroke a systematic review and meta-analysis. BMC Nephrol. 2018;19:283.

[71] Geulayov G, Casey D, McDonald KC, Foster P, Pritchard K, Wells C, et al. Incidence of suicide, hospital-presenting non-fatal self-harm, and community-occurring self-harm in England. BMJ Open. 2016;6:e010538.

[72] Bøe T, Ganapathy SM, Jacobsen BK, Lund JI, Mehlum L, Tsai AC, et al. Psychiatric disorders and repetition of deliberate self-harm: a longitudinal cohort study. J Affect Disord. 2022;296:489-496.

[73] Pisetsky EM, Thornton LM, Lichtenstein P, Pedersen NL, Bulik CM. Suicide attempts in women with eating disorders. Int J Eat Disord. 2013;46(7):632-640.

[74] John A, Griffiths C, Kontopantelis E, Webb RT, Baker C, Kapur N, et al. The rise in antidepressant poisoning among young people in England. J Affect Disord. 2021;294:873-881.