



OPEN ACCESS

EDITED BY

Elena Casiraghi,
Università Degli Studi di Milano, Italy

REVIEWED BY

Luca Lalli,
National Cancer Institute Foundation (IRCCS),
Italy
Ekaterina Esenkova,
Johannes Gutenberg University Mainz,
Germany

*CORRESPONDENCE

Georgina Cosma
✉ g.cosma@lboro.ac.uk

RECEIVED 15 November 2024

ACCEPTED 12 March 2025

PUBLISHED 27 March 2025

CITATION

Kousovista R, Cosma G, Abakasanga E,
Akbari A, Zaccardi F, Jun GT, Kiani R and
Gangadharan S (2025) Temporal patterns of
multiple long-term conditions in individuals
with intellectual disability living in Wales: an
unsupervised clustering approach to disease
trajectories.
Front. Digit. Health 7:1528882.
doi: 10.3389/fdgth.2025.1528882

COPYRIGHT

© 2025 Kousovista, Cosma, Abakasanga,
Akbari, Zaccardi, Jun, Kiani and Gangadharan.
This is an open-access article distributed
under the terms of the [Creative Commons
Attribution License \(CC BY\)](#). The use,
distribution or reproduction in other forums is
permitted, provided the original author(s) and
the copyright owner(s) are credited and that
the original publication in this journal is cited,
in accordance with accepted academic
practice. No use, distribution or reproduction
is permitted which does not comply with
these terms.

Temporal patterns of multiple long-term conditions in individuals with intellectual disability living in Wales: an unsupervised clustering approach to disease trajectories

Rania Kousovista¹, Georgina Cosma^{1*}, Emeka Abakasanga¹,
Ashley Akbari², Francesco Zaccardi³, Gyuchan Thomas Jun⁴,
Reza Kiani⁵ and Satheesh Gangadharan⁵

¹Computer Science Department, School of Science, Loughborough University, Loughborough, United Kingdom, ²Faculty of Medicine, Health and Life Science, Swansea University, Swansea, United Kingdom, ³Leicester Diabetes Center, University of Leicester, United Kingdom, ⁴School of Design and Creative Arts, Loughborough University, Loughborough, Leicester, United Kingdom, ⁵Learning Disability Service (Agnes Unit), Leicestershire Partnership NHS Trust, Leicester, United Kingdom

Introduction: Identifying and understanding the co-occurrence of multiple long-term conditions (MLTCs) in individuals with intellectual disability (ID) is crucial for effective healthcare management. Individuals with ID often experience earlier onset and higher prevalence of MLTCs compared to the general population, however, the specific patterns of co-occurrence and temporal progression of these conditions remain largely unexplored. This study presents an innovative unsupervised approach for examining and characterising clusters of MLTC in individuals with ID, based on their shared disease trajectories.

Methods: Using a dataset of electronic health records (EHRs) from 13,069 individuals with ID, encompassing primary and secondary care data in Wales from 2000 to 2021, this study analysed the time sequences of disease diagnoses. Significant pairwise disease associations were identified, and their temporal directionality assessed. Subsequently, an unsupervised clustering algorithm—spectral clustering—was applied to the shared disease trajectories, grouping them based on common temporal patterns.

Results: The study population comprised 52.3% males and 47.7% females, with a mean of 4.5 ± 3 long-term conditions (LTCs) per patient. Distinct MLTC clusters were identified in both males and females, stratified by age groups (<45 and ≥ 45 years). For males under 45, a single cluster dominated by neurological conditions (32.4%), while three clusters were identified for older males, with the largest characterised by circulatory (51.8%). In females under 45, one cluster was found with digestive system conditions (24.6%) being most prevalent. For females ≥ 45 years, two clusters were identified: the first cluster was predominantly defined by circulatory (34.1%), while the second cluster by digestive (25.9%) and musculoskeletal (21.9%) system conditions. Mental illness, epilepsy, and reflux disorders were prevalent across all groups.

Discussion: This study reveals complex multimorbidity patterns in individuals with ID, highlighting age and sex differences. The identified clusters provide

new insights into disease progression and co-occurrence in this population. These findings can inform the development of targeted interventions and risk stratification strategies, potentially improving personalised healthcare for individuals with ID and MLTCs with the aim of improving health outcome for this vulnerable group of patients i.e. reducing frequency and length of hospital admissions and premature mortality.

KEYWORDS

disease trajectories, chronic disease, co-morbidity, clustering, intellectual disability, multimorbidity

Introduction

People with intellectual disability (ID) face a significantly higher risk of developing a range of physical and mental health conditions compared to the general population. These conditions often occur at a younger age and lead to poorer outcomes, owing to a combination of genetic, behavioural, and social factors (1–3). Studies show a much higher occurrence of multiple long-term conditions (MLTCs) in this population (2). MLTCs, defined as two or more conditions in addition to ID, is linked to premature death and poorer quality of life (4). Despite this, there appear to be only a few studies reporting the prevalence of MLTCs conducted on a large scale (2, 5), but no studies were found to reveal patterns of MLTCs and conditions more likely to co-occur together in this population.

The growing use of electronic health records (EHRs) has enabled significant advances in addressing clinical challenges, enhancing diagnostic capabilities, and improving patient outcomes (6–8). In addition to enabling studies on co-occurring conditions, the longitudinal nature of EHR provides a unique opportunity to uncover temporal associations and trajectories between conditions. Importantly, chronic health conditions frequently co-occur more than expected by chance, often as a consequence of shared risk factors, pathogenicity, or their treatment (9). However, most prior studies have not incorporated the time dimension due to the short time span of the available data (10, 11).

Only recently have a few large-scale analyses assessed disease trajectories by evaluating temporal ordering of co-morbidity pairs over time in general population (12–14). Many studies further developed the framework initially proposed by Jensen et al. (12), who described general principles for temporal trajectory analysis using Danish national data. For instance, Siggaard et al. (15) published a browser of these results, while Jørgensen and Brunak (16) focused on chronic obstructive pulmonary disease (COPD) trajectories. Hu et al. (17) linked the data to a cancer registry to investigate pre-diagnosis trajectories. Jensen et al.'s (12) approach has been applied, with modifications, to other populations including post-depression trajectories in UK Biobank (18), and end-of-life trajectories in California (19). Furthermore, Giannoula et al. (13, 20) proposed a framework to detect and cluster co-morbidity pairs and shared trajectories over time using a dynamic time warping (DTW)-based unsupervised algorithm in

EHRs, and later extended this to include genetic information in the clustering step. Unsupervised algorithms discover natural patterns in data without learning predefined outcomes or classifications. Trajectory analyses can reveal complex, time-ordered condition associations, as well as MLTCs patterns to enable better understanding of disease progression for improved prediction outcomes.

In this study, we propose a computational framework for the analysis of temporal MLTCs on EHRs in 13,069 adults diagnosed with ID in Wales, incorporating 40 long-term conditions (LTCs) from both primary and secondary care data. While most prior studies have applied temporal trajectory analysis to secondary care data and ICD-9 or -10 codes (12, 13, 15–20), with only one study using primary care data (21), our approach utilises both to fully capture MLTCs, as most chronic conditions are treated in general practice. This approach highlights several differences in MLTC patterns between male and female sub-populations across different age groups, acknowledging sex and age as crucial factors in understanding MLTCs. The primary contributions of this research include statistical analysis to identify significant temporal condition pairs, identification of shared MLTCs trajectories, construction of a network of all shared trajectories, and identification of trajectory clusters using an unsupervised machine learning algorithm.

Materials and methods

Study design and participants

This study focused on Welsh residents aged 18 and older with intellectual disability and at least one long-term condition (LTC) between 1st January 2000, and 31st December 2021. This population-based study utilised the Secure Anonymised Information Linkage (SAIL) Databank, a Welsh data repository that enables individual-based data linkage across datasets (22). We identified eligible individuals who were registered with a general practitioner (GP) at the study start date (Supplementary Figure S1). For inclusion in the cohort, individuals required key identifying information as defined within Wales, including a unique anonymised patient identifier, age (or date of birth), sex, residential (WIMD) and GP registration information. Primary and secondary care electronic health records (EHRs) securely

stored within SAIL were used to capture LTC diagnoses. Demographic data were used from the Welsh Demographic Service Dataset (WDSD) that contains information relating to people who are resident in Wales and registered with a Welsh GP. Data collected by GPs is captured via Read v2 codes (5-digit codes related to diagnosis, medication, and process of care codes). Hospital in- and out-patient data are collected in the Patient Episode Database for Wales, which contains clinical information regarding patients' hospital admissions, discharges, diagnoses and operations utilising the International Classification of Diseases (ICD-10) clinical coding system. The Annual District Death Extract (ADDE) from the Office for National Statistics (ONS) was used to capture all deaths and dates of death that occurred over the study period for all Welsh residents, contains information regarding the dates and causes of deaths (also ICD-10).

In this study, a LTC is defined as a condition that cannot, at present, be cured but is controlled by medication and/or other treatment/therapies (23). For conditions that do not always fall into the chronic category, we applied duration-based criteria to define them as long-term or chronic (Supplementary Figure S2). For the purpose of this study, multiple long-term conditions (MLTCs) were defined as two or more chronic conditions. We selected 40 LTCs for this study (Supplementary Table S1), based on consensus from a multidisciplinary professional advisory panel. The professional advisory panel (PAP) comprising a team of experts, including General Practitioners, a consultant Psychiatrist, nurses, pharmacists, and data analysts. The full details of condition merging, grouping, and the comprehensive list of Read v2 and ICD-10 codes for each condition can be found in our study protocol (24).

Data analysis and computational methods

We introduced a novel methodology for identifying and analysing shared disease trajectories in patients with ID. Our approach comprises three main stages: (1) identify pairwise condition associations and their temporal directions; (2) construct shared MLTC trajectories; and (3) utilise a network-based technique to cluster these trajectories into meaningful clusters of similar disease trajectories.

Identifying the pairwise condition associations and their temporal direction

The extracted primary (i.e., GP) and secondary (i.e., hospital) data for all patients were harmonised to a single table, where each row represents a unique patient and each unique variable column represents a binary indicator of the patient's diagnosis of one of the LTCs. For every patient, we extracted the date of first diagnosis for each LTC, creating a chronological sequence of LTC diagnoses. The analysis was stratified according to sex and age. Age was categorised into two groups, under 45 years old (<45) and 45 years old and above (≥ 45), taking into account the median observed age per patient's clinical history. The age stratification threshold of 45 years was selected based on epidemiological literature indicating significant increases in

chronic condition prevalence after this age (25, 26), and evidence that individuals with intellectual disabilities experience earlier onset of aging-related health conditions around this age (27).

Thereafter, we derived all possible pairwise combinations of LTCs from our dataset, considering only those where at least 10 patients shared both conditions and had a minimum temporal separation of six months between the diagnoses. The threshold of at least 10 patients sharing both conditions was selected to provide statistical robustness and minimise the risk of identifying incidental associations without clinical meaning, consistent with similar multimorbidity studies (12, 13, 15–17, 20, 28). The minimum six-month interval between diagnoses was established after consultation with our professional advisory panel to capture temporal progression rather than simultaneous diagnoses. The six-month separation criterion was established to distinguish true disease progression from conditions diagnosed during a single clinical evaluation period, allowing us to identify meaningful temporal relationships. Fisher's exact test was then implemented on 2×2 contingency tables constructed for each qualifying pair of conditions. The resulting p-values were then corrected using the Bonferroni correction for multiple testing, with a threshold of $\alpha = 0.001$. For all co-morbidity pairs that demonstrated a significant association, we assessed whether a statistically significant temporal order (direction) existed between condition 1 (C_1) and condition 2 (C_2). Specifically, a Binomial test was used to evaluate the temporal direction of diagnoses, comparing the number of patients for whom condition C_2 follows condition C_1 against those where C_1 follows C_2 ($C_1 \rightarrow C_2$ vs. $C_2 \rightarrow C_1$). If the p-value < 0.05 from the Binomial test, a preferred direction was assigned to that pair of conditions, based on the more frequently occurring sequence.

Identifying the shared MLTC trajectories

Shared disease trajectories were developed by combining significant temporally ordered condition pairs into longer sequences of MLTCs. These pairs were combined to form all possible longer trajectories. For example, if the pairs $C_1 \rightarrow C_2$ and $C_2 \rightarrow C_3$ were found to be significant, they were combined to form the trajectory $C_1 \rightarrow C_2 \rightarrow C_3$. Pairs with statistically significant (preferred) directionality were included (p-values < 0.05 in the Binomial test for directionality), while in the case of no preferred directionality (p-values ≥ 0.05), both directions were considered. The actual occurrences of these trajectories in the patient population were then counted. These trajectories could contain other intermediate conditions, as long as the conditions maintained their significant chronological order. Consequently, a list of all identified trajectories was obtained, along with their respective occurrence counts in the patient cohort. These trajectories varied in length, with the longer ones representing sequences of conditions where two or more patients shared the exact same chronological order. All trajectories with a length of three conditions shared in more than ten patients used for clustering. For trajectories containing the same three conditions, only the most frequent unique sequence was retained, ensuring each set of three conditions was represented by its most common temporal order.

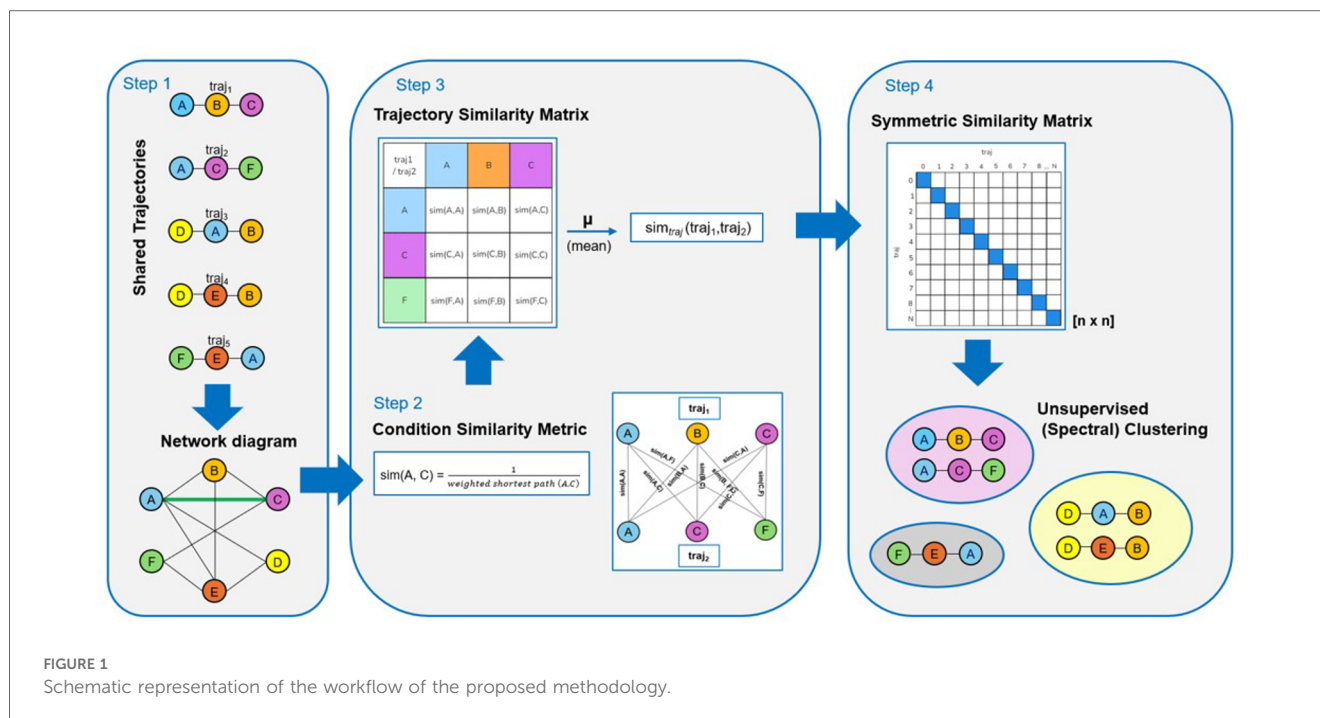


FIGURE 1
Schematic representation of the workflow of the proposed methodology.

Identifying trajectory clusters

We propose a network-based clustering technique that employs a novel approach to quantify and analyse the associations between conditions across multiple disease trajectories. This method consists of four key steps: constructing a trajectory condition network, developing a condition similarity metric based on shortest paths, creating a trajectory similarity matrix, and applying spectral clustering. These steps are described as follows.

Step 1. The first step is about constructing a *trajectory condition network*. A network of all trajectories was constructed to explore the associations between conditions across multiple trajectories and define a similarity metric among them (Figure 1). Let $G = (N, E)$ be an undirected graph (or network graph) where N is the set of nodes (conditions) and E is the set of edges. Each edge $e_{ij} \in E$ connects two conditions c_i and c_j , where $i, j \in 1, 2, \dots, k$, where k is the total number of unique conditions across all trajectories.

Step 2. To quantify similarity between trajectory conditions we developed a *condition similarity metric* based on the shortest path problem in graph theory. The weight w of an edge e_{ij} , where $w: E \rightarrow \mathbb{R}$, is defined as shown in Equation 1:

$$w(e_{ij}) = \frac{1}{\sqrt{f(e_{ij})}} \quad (1)$$

where $f(e_{ij})$, $f: E \rightarrow \mathbb{R}$, is the frequency of edges between conditions c_i and c_j in the network. This weight function assigns lower weights to more frequent edges, effectively making frequently co-occurring conditions closer in the network graph.

Let $P = (c_1, c_2, \dots, c_n)$ be a path from c_1 to c_n in network

graph G . The shortest path length is calculated using Equation 2:

$$\text{shortestpath}(c_1, c_n) = \min_p \sum_{l=1}^{n-1} w(e_{l,l+1}) \quad (2)$$

where the minimum is taken over all possible paths from c_1 to c_n in the graph, and $w(e_{l,l+1})$ is the weight of the edge between consecutive conditions c_l and c_{l+1} in the path. We compute this shortest path using Dijkstra's algorithm (29). To convert the shortest path lengths into similarity scores ranging from 0 to 1, we take the inverse as shown in Equation 3:

$$\text{sim}(c_i, c_j) = \frac{1}{\text{shortest path}(c_i, c_j)} \quad (3)$$

This transformation ensures that conditions with shorter path lengths between them (i.e., more closely related) have higher similarity scores.

Step 3. To assess the overall similarity between trajectories, we constructed an $n \times n$ *symmetric similarity matrix*, where n is the total number of trajectories (Figure 1). The similarity between two trajectories, traj_i and traj_j , is calculated as the mean of the shortest path similarities between their respective condition, as formulated in Equation 4:

$$\text{sim}_{\text{traj}}(\text{traj}_i, \text{traj}_j) = \frac{1}{|\text{traj}_i| \cdot |\text{traj}_j|} \sum_{c_i \in \text{traj}_i} \sum_{c_j \in \text{traj}_j} \text{sim}(c_i, c_j) \quad (4)$$

where $|\text{traj}_i|$ and $|\text{traj}_j|$ are the numbers of conditions in trajectories i and j respectively, $i, j \in 1, 2, \dots, n$ where n is the

total number of trajectories, and $sim(c_i, c_j)$ is the shortest path similarity between conditions c_i and c_j as defined in Equation 3.

Step 4. To identify clusters of highly similar trajectories based on their condition similarities, we applied spectral clustering to the precomputed trajectory similarity matrix (sim_{traj}). We employed spectral clustering for trajectory analysis due to its proven efficacy with complex network structures. Unlike traditional clustering methods that assume spherical or convex cluster shapes, spectral clustering operates by transforming the similarity matrix into a lower-dimensional space using eigenvectors of the Laplacian matrix, making it particularly suitable for our disease trajectory network where relationships are defined by shortest-path distances. This graph-theoretic approach (30) enables detection of meaningful patterns in multimorbidity progression, even when trajectories exhibit nonlinear or overlapping relationships. We implemented spectral clustering using the Scikit-Learn library with default parameter settings (31). The algorithm treats the trajectory similarity matrix as a weighted graph adjacency matrix, performs a spectral embedding of the data points into a lower-dimensional space, and then clusters the embedded points using the k-means algorithm. The optimal number of clusters was determined using the Calinski-Harabasz score.

Results

Data analysis

Table 1 summarises the counts and percentages of patient demographic characteristics included in the study and reports the mean number of LTCs, prevalence of MLTCs (%), and prevalence of physical-mental MLTCs (%) stratified by sex, age group, ethnic group, and Welsh Index of Multiple Deprivation (WIMD) quintiles. The study population comprised 13,069 patients, with 52.3% being male. The mean number of LTCs for all patients was $4.5 (\pm 3)$, with 85.9% having MLTCs and 31.8% having physical-mental MLTCs. Subgroup analysis revealed notable trends across demographic categories. Females showed a slightly higher mean number of LTCs (4.9 ± 3.3) and prevalence of MLTCs (87.7%) compared to males (4.2 ± 2.8 and 84.3% respectively) (Tables 1, Supplementary Table S2). Patients aged 45 years and older had a higher mean number of LTCs (5.2 ± 3.2) and prevalence of MLTCs (91%) compared to those under 45 (3.5 ± 2.4 and 79%). Socioeconomic factors, indicated patients from the most deprived areas had a higher mean number of LTCs (4.9 ± 3.2) and prevalence of MLTCs (88.9%) compared to those from the least deprived areas (4.5 ± 2.9 and 85.6%).

Analysis of pairwise condition associations

Supplementary Table S3 shows the total number of patients after stratification. A comparison of the prevalence of several conditions among males and females, as well as between age groups, can be found in Supplementary Tables S4, S5. The statistically significant condition pairs are provided in Table 2.

TABLE 1 Patient characteristics and description of LTC.

Group	Patients, N (%)	Mean LTC ($\pm$ SD)	Patients with MLTC (%) ^a	Patients with physical-mental MLTC (%) ^a
All patients	13069 (100.0)	4.5 (3.0)	11231 (85.9)	4162 (31.8)
Sex				
Male	6830 (52.3)	4.2 (2.8)	5757 (84.3)	2055 (30.1)
Female	6239 (47.7)	4.9 (3.3)	5474 (87.7)	2107 (33.8)
Age (years)				
< 45	5506 (42.1)	3.5 (2.4)	4350 (79.0)	1800 (32.7)
≥ 45	7563 (57.9)	5.2 (3.2)	6881 (91.0)	2362 (31.2)
Ethnic group				
White	9161 (70.1)	4.8 (3.1)	8074 (88.1)	3259 (35.6)
Unknown	3638 (27.8)	3.8 (2.7)	2916 (80.2)	816 (22.4)
Asian	179 (1.4)	4.6 (2.7)	167 (93.3)	60 (33.5)
Black	43 (0.3)	3.8 (2.9)	35 (81.4)	14 (32.6)
Mixed	15 (0.1)	3.7 (2.7)	12 (80.0)	5 (33.3)
Other	33 (0.3)	4.1 (3.1)	27 (81.8)	8 (24.2)
WIMD				
1. Most deprived	3195 (24.4)	4.9 (3.2)	2840 (88.9)	1148 (35.9)
2	2682 (19.8)	4.6 (3.1)	2232 (86.4)	834 (32.3)
3	2567 (19.6)	4.4 (3.0)	1797 (86.1)	597 (28.6)
4	1949 (14.9)	4.5 (2.9)	1699 (87.2)	564 (28.9)
5. Least deprived	1346 (10.3)	4.5 (2.9)	1152 (85.6)	379 (28.2)

^aProportions calculated using the corresponding subgroup number in "Patients, N(%)" column as the denominator. LTC, long-term condition; SD, standard deviation; MLTC, multiple long-term condition; Physical-mental MLTC, having at least one physical and one mental health condition; WIMD, welsh index of multiple deprivation quintiles categories range from 1 (most deprived) to 5 (least deprived).

Males

Mental illness was the most prevalent condition affecting 32.8% of all males (Supplementary Table S4). A key finding was that its prevalence was higher in those under 45 years (35%) compared to those 45 and above (31%). Epilepsy followed closely, present in 31% of all males, again with a higher prevalence in the younger group (35.3% vs. 27.7%). Reflux disorders showed a consistent prevalence of 29.7% across both age groups. Hypertension demonstrated a marked increase with age, affecting 15.5% of males under 45 and 29.7% of those 45 and above. Additionally, chronic kidney disease (CKD) showed a significant increase from 10.4% in the younger group to 26.9% in the older group. Similarly, diabetes increased from 13.8% in the younger group to 25.3% in the older group.

The most frequent condition pair association in males under 45 (encountered in 385 patients) was found to be between mental illness and reflux disorders (Table 2). Other significant comorbidities include pairs of different neuropsychiatric conditions (e.g., epilepsy $\rightarrow$ mental illness in 287 patients, mental illness $\rightarrow$ insomnia in 267 patients, and cerebral palsy $\leftrightarrow$ epilepsy in 205 patients), combinations of neuropsychiatric and haematological conditions (e.g., mental illness $\rightarrow$ anaemia, 267 patients) or associations between endocrine and circulatory with renal conditions (e.g., diabetes $\rightarrow$ CKD and hypertension $\leftrightarrow$ CKD in

TABLE 2 Top co-morbidity pairs in males and females by age group.

Condition pairs	N patients	Mean years (SD)	Condition pairs	N patients	Mean years (SD)
Males < 45 years			Males ≥ 45 years		
Mental illness → reflux disorders	385	6.7 (4.9)	Hypertension → CKD	435	7.2 (4.9)
Epilepsy → mental illness	287	7.6 (6.4)	Mental illness ↔ reflux disorders	422	7.2 (5.9)
Mental illness → insomnia	267	5.9 (4.8)	Hypertension → diabetes	411	5.7 (4.5)
Insomnia ↔ reflux disorders	216	4.3 (4.1)	Diabetes → CKD	381	5.6 (4.6)
Cerebral palsy ↔ epilepsy	205	2.7 (4.1)	Reflux disorders ↔ chr. arthritis	354	6.4 (5.3)
Chr. airway diseases → insomnia	167	6.9 (4.6)	Chr. airway diseases → reflux disorders	352	5.8 (5.1)
Diabetes ↔ hypertension	149	3.8 (3.6)	Hypertension → chr. arthritis	340	6.9 (4.9)
Mental illness ↔ chr. pain conditions	139	6.4 (4.8)	CKD ↔ anaemia	331	3.0 (3.4)
Reflux disorders ↔ IBD	134	5.4 (4.5)	Cardiac arrhythmias ↔ CKD	318	3.2 (3.9)
Epilepsy → hypertension	121	9.5 (6.1)	Chr. airway diseases → chr. arthritis	291	6.4 (5.0)
Reflux disorders ↔ chr. pain conditions	121	5.0 (4.0)	Reflux disorders → anaemia	285	6.9 (5.4)
Dysphagia ↔ reflux disorders	118	3.1 (3.5)	Hypertension → cardiac arrhythmias	279	7.5 (5.3)
Reflux disorders → anaemia	107	5.5 (4.5)	Epilepsy → dysphagia	256	9.5 (7.1)
Epilepsy → chr. constipation	95	6.1 (5.0)	Hypertension → CHD	256	6.2 (4.6)
Insomnia → chr. arthritis	92	6.1 (5.0)	Cardiac arrhythmias → heart failure	253	1.7 (3.0)
Diabetes → CKD	82	7.5 (6.8)	Chr. airway diseases → cardiac arrhythmias	245	7.2 (5.6)
Hypertension ↔ CKD	82	7.4 (4.6)	CHD → CKD	244	5.1 (4.5)
Chr. pain conditions → neuropathic pain	77	4.2 (5.2)	Chr. airway diseases → CHD	243	5.3 (5.3)
Chr. airway diseases → cardiac arrhythmias	76	7.8 (5.6)	Chr. arthritis → cardiac arrhythmias	242	7.1 (5.6)
Cerebral palsy → dysphagia	72	8.5 (6.5)	Diabetes → cardiac arrhythmias	239	5.9 (5.0)
Females < 45 years			Females ≥ 45 years		
Mental illness → reflux disorders	375	7.4 (5.9)	Hypertension → chr. arthritis	484	6.3 (5.1)
Chr. airway diseases → mental illness	357	7.9 (6.0)	Mental illness ↔ reflux disorders	473	6.2 (5.5)
Chr. airway diseases → reflux disorders	293	6.5 (5.3)	Mental illness → chr. arthritis	470	7.2 (5.7)
Chr. pain conditions ↔ mental illness	260	5.7 (5.1)	Hypertension → CKD	462	7.0 (5.1)
Chr. pain conditions → reflux disorders	252	7.3 (5.2)	Chr. arthritis ↔ reflux disorders	460	5.6 (5.1)
Mental illness → insomnia	240	6.2 (5.1)	Mental illness ↔ chr. airway diseases	429	7.5 (6.2)
Epilepsy → mental illness	229	7.3 (5.9)	Chr. arthritis → CKD	405	6.9 (5.2)
Chr. airway diseases → chr. pain conditions	220	7.8 (5.9)	Diabetes ↔ hypertension	405	4.8 (4.2)
Reflux disorders ↔ anaemia	199	5.6 (5.0)	Chr. airway diseases → chr. arthritis	391	6.2 (5.2)
Reflux disorders ↔ insomnia	179	5.5 (3.8)	Chr. airway diseases → reflux disorders	386	6.2 (5.1)
Mental illness → diabetes	167	6.8 (5.8)	Menopausal & perimenopausal ↔ chr. arthritis	365	5.8 (5.0)
Chr. airway diseases → insomnia	164	8.7 (5.8)	Diabetes → CKD	361	5.7 (4.5)
Epilepsy → thyroid disorders	160	7.4 (5.3)	Mental illness → diabetes	359	7.8 (6.0)
Epilepsy ↔ cerebral palsy	158	8.0 (6.1)	Reflux disorders → anaemia	338	5.1 (4.6)
Mental illness → chr. arthritis	154	8.9 (6.3)	Chr. arthritis → anaemia	328	5.9 (5.2)
Mental illness → IBD	152	7.7 (4.7)	Chr. airway diseases → diabetes	300	5.5 (5.1)
Mental illness → neuropathic pain	147	7.2 (5.3)	Thyroid disorders → Dementia	286	8.5 (5.8)
Chr. airway diseases → diabetes	136	7.8 (5.6)	Chr. arthritis → cardiac arrhythmias	278	6.7 (5.5)
Chr. airway diseases → chr. arthritis	132	7.6 (5.9)	Hypertension → cardiac arrhythmias	267	8.2 (5.6)
Epilepsy → dysphagia	130	9.5 (6.3)	Diabetes → anaemia	267	6.1 (4.2)

Arrows indicate temporal associations: → suggests the first condition typically precedes the second, while ↔ indicates no preferred order. The total number of patients (N Patients) represents the number of individuals with each co-morbidity pair, regardless of order. Mean Years ± SD shows the average time between the occurrences of the two conditions, regardless of order. Chr., chronic; CKD, chronic kidney disease; CHD, coronary heart disease.

82 patients). In males aged 45 and above, hypertension → CKD emerged as the most prevalent pair, affecting 447 patients, followed by mental illness ↔ reflux disorders (422 patients) and hypertension → diabetes (411 patients).

Females

Mental illness was also the most common condition among females, affecting 35.1% overall (Supplementary Table S5). Thyroid conditions were prevalent in 30.8% of all females, with 33.3% in the 45 and above age group and 27.1% in those under

45. Reflux disorders affected 30.1% of all females, with 30.9% in the under 45 group and 29.5% in those 45 and above. Hypertension showed a significant age-related increase, from 11.9% in those under 45 to 31.9% in those 45 and above. As expected, menopausal and perimenopausal conditions were more prevalent in the older female group, affecting 25.9% of those 45 and above compared to 8.8% in those under 45.

The most frequent condition pair in females under 45 was mental illness → reflux disorders, shared by 375 patients (Table 2). Chronic airway conditions featured prominently in

TABLE 3 Long-term condition clusters for (A) males and (B) females categorised by age groups. Each cluster presents the most frequent system condition categories included in the trajectories alongside the count of trajectories (*N* traj) and total patient numbers (*N* patients). System percentages (%) are calculated based on the total number of trajectories in each cluster. Clusters are presented in descending order of patient count within each age category.

Cluster	<i>N</i> traj	<i>N</i> patients	System condition groups distribution (%)	Mortality % ^a	Long hospital stay % ^{a,b}
(A) Males					
(a) < 45 years					
Cluster 1	37	549	Nervous (32.4%), musculoskeletal (19.8%), digestive (18.0%), mental (12.6%), respiratory (7.2%), circulatory (3.7%), endocrine (2.7%), genitourinary (1.8%), blood (1.8%)	16.9	46.8
(b) ≥ 45 years					
Cluster 1	112	2824	Circulatory (51.8%), genitourinary (11.9%), endocrine (8.0%), blood (7.1%), respiratory (7.1%), musculoskeletal (4.8%), nervous (4.8%), digestive (4.5%)	63.5	84.1
Cluster 2	81	1557	Musculoskeletal (25.1%), digestive (20.6%), nervous (18.1%), circulatory (14.4%), respiratory (8.6%), blood (5.3%), mental (4.5%), ear (3.4%)	47.8	67.4
Cluster 3	36	633	Digestive (39.8%), mental (15.7%), nervous (13.9%), respiratory (13.9%), circulatory (5.6%), genitourinary (4.6%), blood (2.8%), ear (1.9%), endocrine (1.9%)	69.4	82.2
(B) Females					
(a) < 45 years					
Cluster 1	88	1713	Digestive (24.6%), nervous (21.6%), musculoskeletal (18.6%), mental (14.4%), respiratory (10.2%), endocrine (5.3%), blood (2.3%), circulatory (1.5%), genitourinary (1.5%)	13.4	43.8
(b) ≥ 45 years					
Cluster 1	256	6101	Circulatory (34.1%), nervous (12.1%), musculoskeletal (10.5%), genitourinary (8.9%), endocrine (8.5%), respiratory (6.9%), blood (6.6%), mental (5.5%), digestive (3.9%), neoplasms (2.2%), ear (0.8%)	58.5	74.9
Cluster 2	183	4057	Digestive (25.9%), musculoskeletal (21.9%), nervous (16.2%), genitourinary (7.8%), circulatory (7.7%), respiratory (6.7%), mental (6.6%), blood (4.0%), ear (1.6%), endocrine (1.5%), neoplasms (0.2%)	43.0	61.8

^aPercentages (%) are calculated using the unique number of patients included in each cluster.

^bLong hospital stay is defined as a hospitalisation lasting more than 4 days.

this group, with chronic airway diseases → mental illness and chronic airway diseases → reflux disorders affecting 357 and 293 patients, respectively. In the older age groups, we noted significant associations between circulatory and musculoskeletal conditions (e.g., hypertension → chronic arthritis, 484 patients), followed by mental illness → reflux disorders (473 patients) and mental illness → chronic arthritis (470 patients).

In general, the younger age groups, particularly those under 45, showed a higher prevalence of co-morbidities involving neuropsychiatric conditions, often in combination with other system conditions (Table 2). This was evident in both males and females, with mental illness frequently co-occurring with conditions from various other physiological systems. While mental illness and epilepsy were among the top five most prevalent conditions in both sexes, thyroid disorders ranked second in females but did not appear in the top five for males (Supplementary Tables S4, S5). Hypertension was among the top five conditions in males across all ages, but in females, it only ranked in the top five for the 45 and above age group.

Analysis of trajectory clusters

Identified clusters in trajectories of males

In the male sub-population, one cluster was identified for males under 45 years (<45) and three clusters for males 45 years and older (≥45) (Supplementary Table S6). Thirty-seven shared trajectories were considered for clustering in males <45 years,

and 229 shared trajectories in males ≥45 years (Supplementary Table S7). Only trajectories with a minimum of ten patients were included. Table 3 presents an overview of the clusters identified within the male population.

For males under 45 years

The analysis of shared trajectories in males revealed distinct patterns across different age groups (Figure 2). For males under 45 years, neurological conditions were the most common affecting 32.4% of this subpopulation. This was followed by conditions of the musculoskeletal (19.8%) and digestive system (18.0%). Mental health conditions were also notable, present in 12.6% of cases. Looking at more specific conditions of this younger cohort revealed an equal prevalence (37.8%) of chronic pain conditions, mental illness, and insomnia. Other significant conditions included reflux disorders (35.1%), epilepsy (27.0%), and chronic airway diseases (21.6%) (Supplementary Tables S8, S10).

For males 45 years and older

The analysis of health trajectories in males aged 45 years and older revealed three distinct clusters, each characterised by unique patterns of multimorbidity. The largest cluster, encompassing 112 shared trajectories and 2,824 patients, was predominantly defined by circulatory system conditions (51.8%) (Table 3). Coronary heart disease (CHD) emerged as the most prevalent condition, affecting 44.6% of patients, closely followed by CKD (35.7%) and cardiac arrhythmias (34.8%) (Supplementary Tables S8, S11). Heart failure was also significantly present, occurring in 32.1% of cases. This cluster highlighted a strong interplay between cardiovascular and renal health, further complicated by metabolic

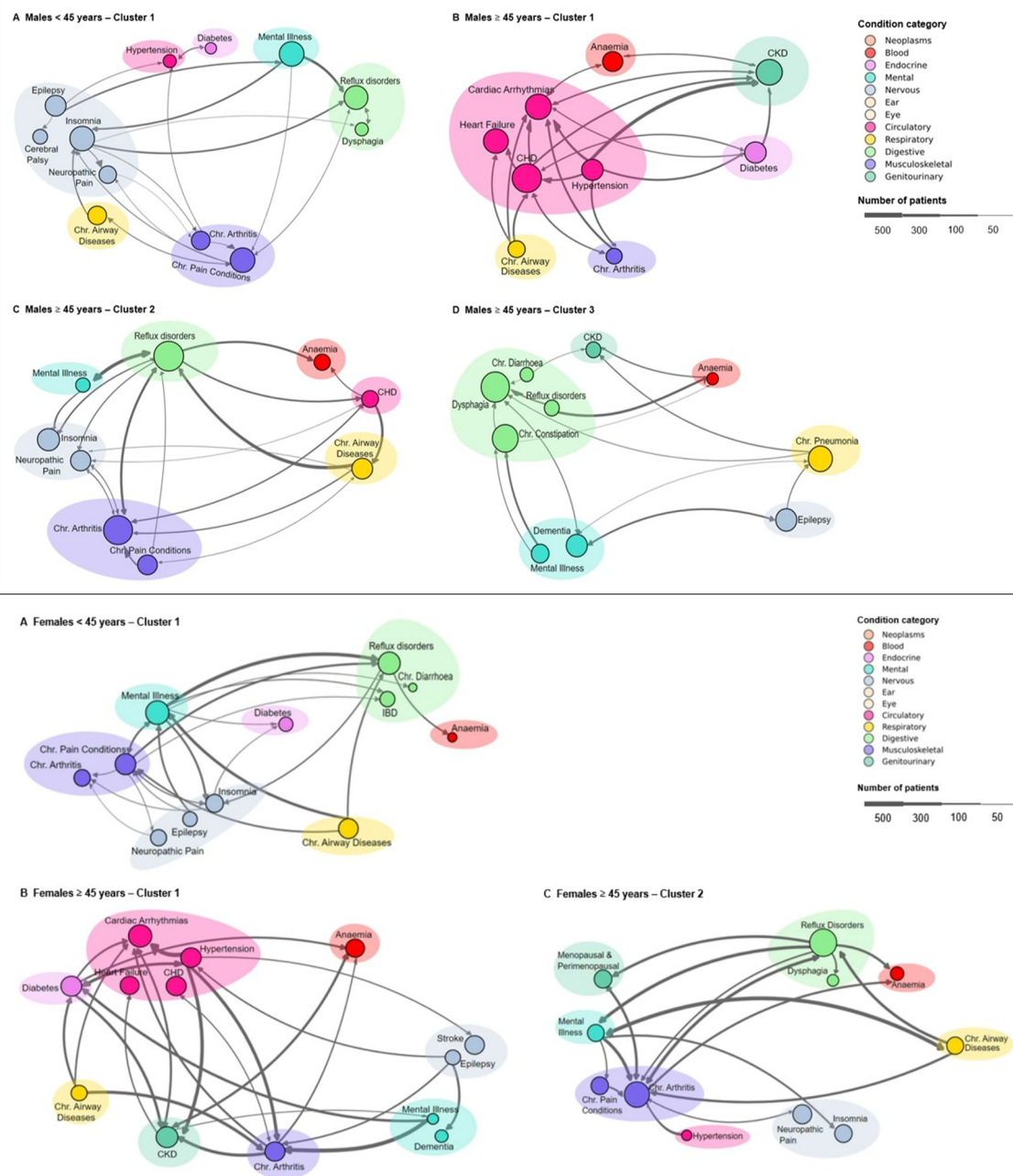


FIGURE 2

Disease trajectory clusters across age groups and genders. Top: Males – (A) <45 years cluster 1, (B) ≥45 years cluster 1, (C) ≥45 years cluster 2, and (D) ≥45 years cluster 3. Bottom: Females – (A) <45 years single cluster; (B) ≥45 years cluster 1, and (C) ≥45 years cluster 2. Node size indicates condition prevalence, with larger nodes representing more frequent occurrences. Conditions in each cluster represent more than 5% of the total conditions. Edges show associations between conditions, with edge thickness corresponding to the frequency of condition pair occurrences (minimum edge frequency of 1%). Coloured shaded areas group related conditions within the same category as defined in the legend. The legend specifies condition categories and provides a scale for patient numbers.

conditions such as diabetes and hypertension, each present in 24.1% of patients. The presence of peripheral vascular disease (19.6%) and chronic airway diseases (15.2%) in this cluster underscores the complex associations between cardiovascular, respiratory, and metabolic systems in ageing men.

The second cluster, comprising 81 trajectories and 1,557 patients, was dominated by musculoskeletal (25.1%) and digestive

system (20.6%) conditions. Reflux disorders were most prevalent, affecting 53.1% of patients, closely followed by chronic arthritis at 50.6% (Supplementary Table S8). This cluster was prominently associated with the high incidence of chronic pain conditions (23.5%) and neuropathic pain (24.7%), indicating that chronic pain significantly affects many older men. Cardiovascular conditions remained relevant, with CHD affecting 17.3% of

patients. The presence of mental illness (12.3%) and hearing loss (8.6%) in this cluster points to the diverse health challenges faced by this subgroup.

The third cluster, though smaller with 36 trajectories and 633 patients was characterised by a high prevalence of digestive (39.8%) and neurological (13.9%) conditions. Dysphagia was the most common condition, present in 50.0% of patients, followed closely by chronic constipation at 41.7% ([Supplementary Table S8](#)). Neurological conditions were prominent, with both epilepsy and dementia affecting 27.8% of patients each. The significant presence of chronic pneumonia (36.1%) in this cluster highlights the persistent relevance of respiratory conditions in older males, while the occurrence of CKD (13.9%) further emphasises the complex interplay of multiple organ systems in ageing.

Identified clusters in trajectories of females

In the female sub-population, three clusters were obtained, one for females under 45 years (<45) and two for females 45 years and older (≥ 45) ([Supplementary Table S6](#)). Eighty-eight shared trajectories were considered for clustering in females <45 years, and 439 shared trajectories in females ≥ 45 years ([Supplementary Table S7](#)). [Table 3](#), [Figure 2](#) illustrates the distinct patterns of health trajectory clusters identified within the female population across different age groups.

For females under 45 years

The analysis of shared trajectories in females revealed distinct patterns across different age groups. For females under 45 years, digestive system conditions emerged as the predominant health concern, affecting 24.6% of this subpopulation. This was closely followed by neurological conditions (21.6%) and musculoskeletal conditions (18.6%). Mental health conditions were also prominent, present in 14.4% of cases. Looking at more specific conditions of this younger cohort revealed a high prevalence of mental illness (43.2%), followed by reflux disorders (38.6%), and chronic pain conditions (34.1%). Other significant conditions included chronic airway diseases (30.7%), insomnia (25.0%), and chronic arthritis (21.6%) ([Supplementary Tables S9, S12](#)).

For females 45 years and older

The analysis of health trajectories in females aged 45 years and older revealed two distinct clusters. The larger of these clusters, encompassing 256 shared trajectories and 6,101 patients, was predominantly defined by circulatory system conditions (34.1%) ([Table 3](#)). Cardiac arrhythmias emerged as the most prevalent condition, affecting 30.5% of patients, closely followed by CKD (26.6%) and diabetes (24.2%) ([Supplementary Table S9](#)). Hypertension was also significantly present, occurring in 23.4% of cases. This cluster highlighted a strong interplay between cardiovascular and metabolic health, further complicated by musculoskeletal conditions such as chronic arthritis, present in 22.7% of patients. The presence of CHD (22.3%) and chronic airway diseases (14.5%) in this group underscores the complex associations between cardiovascular, respiratory, and metabolic systems in ageing women.

A distinct health profile emerged in the other cluster, comprising 183 trajectories and 4,057 patients, dominated by

digestive (25.9%) and musculoskeletal (21.9%) conditions. Reflux disorders were most prevalent, affecting 48.1% of patients, closely followed by chronic arthritis at 41.5% ([Supplementary Tables S9, S13](#)). This group was significant for its high incidence of neuropathic pain (23.5%) and chronic pain conditions (20.2%), indicating that chronic pain significantly affects many older women. Menopausal and perimenopausal conditions were also prominent, affecting 20.8% of patients.

Cluster characteristics by sex and age groups

[Figure 3](#), [Supplementary Table S14](#) present the mortality and long hospital stay (≥ 4 days) rates observed across the identified clusters, stratified by sex and age groups. In the younger age group (<45 years), both males and females exhibited single clusters with similar mean ages (35.6 ± 5.8 and 35.9 ± 6.1 years, respectively). Mortality percentages for clusters <45 years were comparably low, with males at 16.9% and females at 13.4%. Long hospital stays were less common in this age group, with rates of 46.8% for males and 43.8% for females, significantly lower than in ≥ 45 year populations.

For the age group ≥ 45 years, we observed more distinct patterns in mortality and long hospital stays. In males, clusters 1 and 3 showed particularly high mortality percentages (63.5% and 69.4%). These clusters also had the highest rates of long hospital stays (84.1% and 82.2%, respectively). Cluster 1 was predominantly characterised by circulatory system conditions (51.8%), while cluster 3 was notable for digestive system conditions (39.8%), suggesting a strong association between these condition categories and both increased mortality risk and extended hospitalisations. Similarly, in females ≥ 45 years, we observed a similar pattern, with the larger cluster (cluster 1) showing a higher mortality rate (3.5 per 100 patient-years) and percentage (58.5%) compared to cluster 2 (2.4 per 100 patient-years and 43.0%). Cluster 1 also had a higher rate of long hospital stays (74.9% vs 61.8% in cluster 2) and was primarily affected by circulatory system conditions (34.1%).

Analysis of causes of death

[Figure 4](#) presents the distribution of the five leading causes of death among individuals with ID in our study population, which we can compare to the findings from the most recent LeDeR (learning from lives and deaths-people with a learning disability and autistic people) annual report (32). Across all groups, circulatory system conditions consistently appear as a major cause of death, particularly prominent in the older age groups (≥ 45 years). Notably, in the <45 years age groups, neoplasms and respiratory system conditions feature more prominently as causes of death compared to the older groups. In the older age groups, we observe some variations between clusters, with cluster 1 for both males and females showing a higher proportion of deaths due to circulatory system conditions compared to other clusters. This aligns with our earlier observation of cluster 1 being characterised by a higher prevalence of circulatory system conditions. These findings underscore the complex interplay between age, sex, and specific health conditions in determining mortality patterns among individuals with ID.

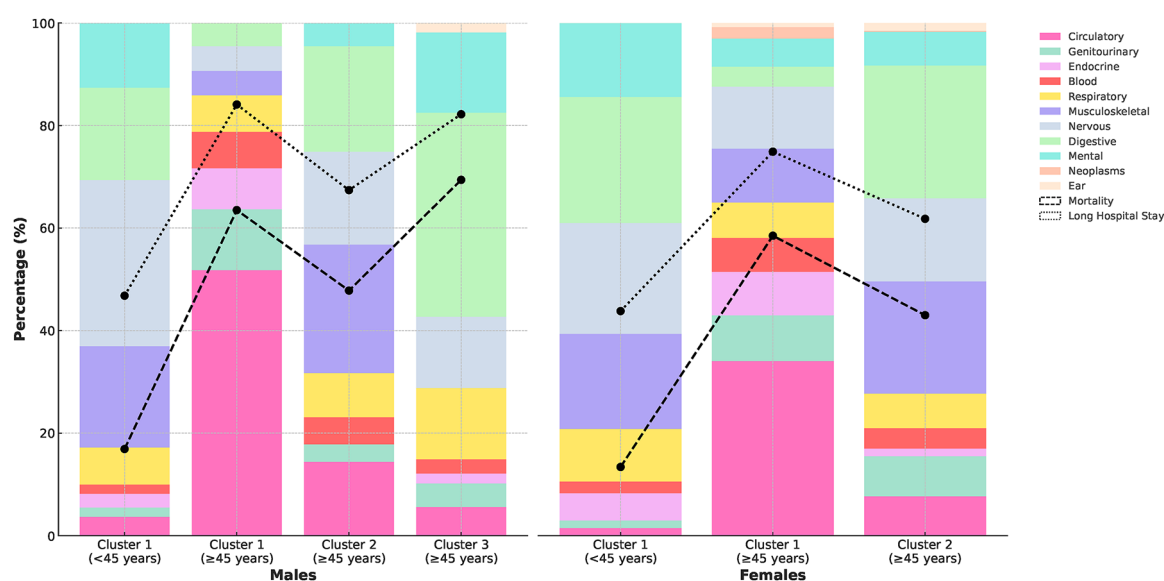


FIGURE 3
Distribution of long-term conditions and rates of mortality and long hospital stays across patient clusters.

Discussion

This study presents a novel approach for the identification and temporal analysis of LTC trajectories in adults with ID using both primary and secondary care data. By employing an unsupervised clustering method on trajectories of a large dataset of electronic health records, we have uncovered distinct patterns of co-morbidity and disease trajectories that vary significantly across sex and age groups. Our findings reveal a high prevalence of MLTCs in the ID population, with 85.9% of individuals experiencing two or more long-term conditions. This aligns with previous research highlighting the complex health needs of this population (1, 33). On average this group has 4.5 long-term conditions per patient.

Sex differences

Our analysis demonstrated significant sex differences in the prevalence of LTCs among adults with ID, while also identifying common patterns. Mental illness and epilepsy were the most prevalent conditions across both sexes, with mental illness affecting more than 30% and epilepsy affecting approximately 30% of both males and females. This prevalence aligns with previous findings in the literature (5). However, clear sex differences were apparent in other conditions. Females exhibited a higher prevalence of thyroid disorders (30.8%) compared to males (14.1%), and anaemia affected 20% of females compared to 14% of males. Females also showed increased prevalence of endocrine, skeletal, and digestive conditions, a finding consistent with previous research (34). Our investigation of temporal co-morbidities revealed that some conditions displayed opposite temporal associations in males and females for certain pairs of conditions. For instance, we observed that in females under 45 years, epilepsy tended to precede thyroid disorders, while this

association was not significant in males of any age group. This finding highlights the importance of investigating sex-stratified disease trajectories, as the temporal direction between conditions can vary depending on sex (28).

Age-related patterns

Age-related differences were apparent in condition patterns between <45 years and ≥45 years adults with ID. Under 45 years groups showed a predominance of neurological (such as epilepsy and dementia), digestive (including dysphagia and chronic constipation), and mental health conditions, consistent with previous research in the ID population (35, 36). In males under 45 years, neurological conditions accounted for 32.4% of the identified cluster, while digestive system conditions were most prevalent (24.6%) in females of the same age group.

In the over 45 groups, circulatory system conditions emerged as a primary concern for both sexes with ID. We found a higher prevalence of CHD, cardiac arrhythmias, heart failure, hypertension, kidney disease, and diabetes in older adults with ID. The co-occurrence of these conditions is particularly noteworthy, as it mirrors the complex interplay of cardiovascular and renal health observed in the general population (37). Furthermore, the strong association between hypertension and various other cardiovascular complications, including heart failure, arrhythmias, and CHD, aligns with patterns seen in the broader population (38).

Mortality and hospitalisation

Across all age groups and clusters, males generally exhibited slightly higher mortality rates and percentages compared to



FIGURE 4

Comparison of the five leading causes of death among individuals with intellectual disabilities (ID), grouped based on ICD-10 chapters. Data presented includes findings from the (A) LeDeR Annual Report 2022 (32), (B) our cohort study using the SAIL (Secure Anonymised Information Linkage) databank and stratification by clusters based on sex and age groups for (C) male cluster (<45 years: cluster 1 and ≥45 years: clusters 1, 2, 3) and (D) females (<45 years: cluster 1; ≥45 years: clusters 1, 2). Values represent percentages for each cause of death.

females—a pattern observed in the general population but appears to be more pronounced in individuals with ID (39). Our analysis revealed a strong association between circulatory system conditions and both increased mortality and extended hospitalisations, particularly in older adults, highlighting these conditions as the most prevalent cause of death in the identified clusters. While mental health and neurological conditions are prevalent across all groups, their impact on mortality appears less direct compared to circulatory and certain digestive system conditions. However, mental conditions significantly contribute to extended hospital care needs and overall quality of life, especially in younger adults with ID (40). We observed a clear association between higher rates of long hospital stays and increased mortality in specific clusters of the older population. For instance, in males ≥ 45 years, cluster 1 exhibited both the highest rate of long stays (84.1%) and a high mortality rate (3.9 per 100 patient-years). Lastly, our study underscores the value of temporality in predicting both co-morbidity and mortality (41).

Limitations

The reliance on clinical records may result in underdiagnosis of certain conditions, particularly those with subtle presentations in the ID population. Moreover, our analysis was constrained by limited ethnic diversity in the dataset. Additionally, the lack of a general population comparison group limits our ability to determine whether observed multimorbidity patterns are unique to individuals with intellectual disabilities or reflect common disease associations. While our clusters demonstrate statistically significant disease groupings and have been reviewed by healthcare professionals, further validation against external clinical benchmarks would strengthen their clinical utility. Future work could involve clinical validation studies to examine how these clusters relate to treatment responses and patient outcomes, potentially transforming these statistical findings into actionable clinical insights.

Conclusion

In conclusion, this study presents a data-driven overview of MLTC trajectory patterns in adults with ID. By revealing distinct clusters of conditions and their progression over time, our findings underscore the complex interplay of age, sex, and health conditions in adults with ID and provide a foundation for more targeted, personalised healthcare strategies. Our work advances understanding of how MLTCs manifest and progress in people with ID, revealing distinct patterns of disease development and complex interactions between conditions in this population. Future research aims to validate these clusters in diverse populations and investigate the underlying mechanisms of the observed progression patterns.

Data availability statement

The datasets presented in this article are not readily available as all proposals to use SAIL data are subject to review by the independent IGRP. The anonymised individual-level data sources used in this study are available in the SAIL Databank at Swansea University, Swansea, UK. Before any data can be accessed, approval must be given by the IGRP. The IGRP gives careful consideration to each project to ensure proper and appropriate use of SAIL data. When access has been granted, it is gained through a privacy-protecting safe haven and remote access system referred to as the SAIL Gateway. SAIL has established an application process to be followed by anyone who would like to access data via SAIL at: <https://www.saildatabank.com/application-process>.

Ethics statement

The studies involving humans were approved by the SAIL Databank's independent Information Governance Review Panel (IGRP Project: 1375), ensuring compliance with ethical standards and data protection regulations. Individual written patient consent was not required for this study. The studies were conducted in accordance with the local legislation and institutional requirements. The ethics committee/institutional review board waived the requirement of written informed consent for participation from the participants or the participants' legal guardians/next of kin because Individual written patient consent was not required for this study.

Author contributions

RK: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Resources,

Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing. GC: Conceptualization, Funding acquisition, Project administration, Resources, Supervision, Writing – review & editing. EA: Data curation, Writing – review & editing. AA: Data curation, Funding acquisition, Resources, Writing – review & editing. FZ: Data curation, Funding acquisition, Writing – review & editing. GTJ: Funding acquisition, Writing – review & editing. RK: Writing – review & editing. SG: Data curation, Funding acquisition, Writing – review & editing, Resources.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. The work was funded by National Institute for Health and Care Research. The project is entitled “DECODE: Data-driven machine-learning aided stratification and management of multiple long-term COnditions in adults with intellectual disabiliTiEs.” Grant no. NIHR203981.

Acknowledgments

We extend our deepest gratitude to our Public and Patient Involvement and Engagement (PPIE) partners including people with intellectual disabilities, family and paid carers. Their invaluable contributions, rooted in their lived personal experience of intellectual disabilities, were instrumental in the research. Without their expertise, insight, and collaborative spirit, this work would not have been possible. We also want to thank Dr. Alison Drewett for her insights regarding the selection of colours for the visualisations. This decision-making supports wider aims pertaining to accessibility for people with intellectual disabilities.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of

their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

- Cooper SA, Allan L, Greenlaw N, McSkimming P, Jasilek A, Henderson A, et al. Rates, causes, place and predictors of mortality in adults with intellectual disabilities with and without down syndrome: cohort study with record linkage. *BMJ Open*. (2020) 10:e036465. doi: 10.1136/bmjopen-2019-036465
- Kinnear D, Morrison J, Allan L, Henderson A, Smiley E, Cooper SA. Prevalence of physical conditions and multimorbidity in a cohort of adults with intellectual disabilities with and without down syndrome: cross-sectional study. *BMJ Open*. (2018) 8:e018292. doi: 10.1136/bmjopen-2017-018292
- Emerson E, Hatton C. *Health Inequalities and People with Intellectual Disabilities*. Cambridge: Cambridge University Press (2014).
- Heslop P, Hoghton M. The learning disabilities mortality review (leder) programme. *Br J Gen Pract*. (2018) 68:bjgp18X697313. doi: 10.3399/bjgp18X697313
- Carey IM, Shah SM, Hosking FJ, DeWilde S, Harris T, Beighton C, et al. Health characteristics and consultation patterns of people with intellectual disability: a cross-sectional database study in english general practice. *Br J Gen Pract*. (2016) 66:e264–70. doi: 10.3399/bjgp16X684301
- Blumenthal D, Tavenner M. The “meaningful use” regulation for electronic health records. *New Engl J Med*. (2010) 363:501–4. doi: 10.1056/NEJMp1006114
- Solares JRA, Raimondi FED, Zhu Y, Rahimian F, Canoy D, Tran J, et al. Deep learning for electronic health records: a comparative review of multiple deep neural architectures. *J Biomed Inform*. (2020) 101:103337. doi: 10.1016/j.jbi.2019.103337
- Pham T, Tran T, Phung D, Venkatesh S. Deepcare: a deep dynamic memory model for predictive medicine. In: *Advances in Knowledge Discovery and Data Mining: 20th Pacific-Asia Conference, PAKDD 2016, Auckland, New Zealand, April 19–22, 2016, Proceedings, Part II* 20. Springer (2016). p. 30–41.
- Johnston MC, Crilly M, Black C, Prescott GJ, Mercer SW. Defining and measuring multimorbidity: a systematic review of systematic reviews. *Eur J Public Health*. (2019) 29:182–9. doi: 10.1093/eurpub/cky098
- Chmiel A, Klimek P, Thurner S. Spreading of diseases through comorbidity networks across life and gender. *New J Phys*. (2014) 16:115013. doi: 10.1088/1367-2630/16/11/115013
- Teno JM, Weitzen S, Fennell ML, Mor V. Dying trajectory in the last year of life: does cancer trajectory fit other diseases? *J Palliat Med*. (2001) 4:457–64. doi: 10.1089/109662101753381593
- Jensen AB, Moseley PL, Oprea TI, Ellesøe SG, Eriksson R, Schmock H, et al. Temporal disease trajectories condensed from population-wide registry data covering 6.2 million patients. *Nat Commun*. (2014) 5:4022. doi: 10.1038/ncomms5022
- Giannoula A, Gutierrez-Sacristán A, Furlong LI. Identifying temporal patterns in patient disease trajectories using dynamic time warping: a population-based study. *Sci Rep*. (2018) 8:4216. doi: 10.1038/s41598-018-22578-1
- Lyons J, Akbari A, Abrams KR, Lorenzo AA, Dhafari TB, Chess J, et al. Trajectories in chronic disease accrual and mortality across the lifespan in Wales, UK (2005–2019), by area deprivation profile: linked electronic health records cohort study on 965,905 individuals. *Lancet Reg Health Eur*. (2023) 32:100687. doi: 10.1016/j.lanepe.2023.100687
- Siggaard T, Reguant R, Jørgensen IF, Haue AD, Lademann M, Aguayo-Orozco A, et al. Disease trajectory browser for exploring temporal, population-wide disease progression patterns in 7.2 million danish patients. *Nat Commun*. (2020) 11:4952. doi: 10.1038/s41467-020-18682-4
- Jørgensen IF, Brunak S. Time-ordered comorbidity correlations identify patients at risk of mis- and overdiagnosis. *NPJ Digit Med*. (2021) 4:12. doi: 10.1038/s41746-021-00382-y
- Hu JX, Helleberg M, Jensen AB, Brunak S, Lundgren J. A large-cohort, longitudinal study determines precancer disease routes across different cancer types. *Cancer Res*. (2019) 79:864–72. doi: 10.1158/0008-5472.CAN-18-1677
- Han X, Hou C, Yang H, Chen W, Ying Z, Hu Y, et al. Disease trajectories and mortality among individuals diagnosed with depression: a community-based cohort study in UK Biobank. *Mol Psychiatry*. (2021) 26:6736–46. doi: 10.1038/s41380-021-01170-6
- Paik H, Kim J. Condensed trajectory of the temporal correlation of diseases and mortality extracted from over 300,000 patients in hospitals. *PLoS One*. (2021) 16:e0257894. doi: 10.1371/journal.pone.0257894
- Giannoula A, Centeno E, Mayer MA, Sanz F, Furlong LI. A system-level analysis of patient disease trajectories based on clinical, phenotypic and molecular similarities. *Bioinformatics*. (2021) 37:1435–43. doi: 10.1093/bioinformatics/btaa964
- Planell-Morell P, Bajekal M, Denaxas S, Raine R, Alexander DC. Trajectories of disease accumulation using electronic health records. *Stud Health Technol Inform*. (2020) 270:469–73. doi: 10.3233/SHTI200204
- Lyons RA, Jones KH, John G, Brooks CJ, Verplanck JP, Ford DV, et al. The sail databank: linking multiple health and social care datasets. *BMC Med Inform Decis Mak*. (2009) 9:1–8. doi: 10.1186/1472-6947-9-3
- Department of Health and Social Care. Long term conditions compendium of information. *Report*. NHS (2012).
- Shabnam S, Kousovista R, Abakasanga E, Kaur N, Cosma G, Akbari A, et al. Data from: Data preparation and epidemiological plan—decode (2024). doi: 10.17605/OSF.IO/KT5FY
- Zheng DD, McCollister KE, Christ SL, Lam BL, Feaster DJ, Lee DJ. Chronic condition patterns in the US population and their association with health related quality of life. *Prev Med*. (2020) 136:106102. doi: 10.1016/j.ypmed.2020.106102
- Buttorff C, Ruder T, Bauman M. Data from: Multiple chronic conditions in the United States (2017). Vol. 10. Santa Monica (CA): RAND.
- García-Domínguez L, Navas P. Chronic health conditions in aging individuals with intellectual disabilities. *Int J Environ Res Public Health*. (2020) 17:3126. doi: 10.3390/ijerph17093126
- Westergaard D, Moseley P, Sørup FKH, Baldi P, Brunak S. Population-wide analysis of differences in disease progression patterns in men and women. *Nat Commun*. (2019) 10:666. doi: 10.1038/s41467-019-08475-9
- Dijkstra EW. A note on two problems in connexion with graphs. In: Hamilton K, Liskov B, Myers BA, editors. *Edsger Wybe Dijkstra: His Life, Work, and Legacy*. New York, NY: Association for Computing Machinery (2022). p. 287–90.
- Ding L, Li C, Jin D, Ding S. Survey of spectral clustering based on graph theory. *Pattern Recognit*. (2024) 147:110366. doi: 10.1016/j.patcog.2024.110366
- Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: machine learning in python. *J Mach Learn Res*. (2011) 12:2825–30. doi: 10.48550/arXiv.1201.0490
- White A, Sheehan R, Ding J, Roberts C, Magill N, Keagan-Bull R, et al. *Learning from Lives and Deaths—People with a Learning Disability and Autistic People (Leder) Report for 2022*. London: King's College London (2023).
- Mann C, Jun GT, Tyrer F, Kiani R, Lewin G, Gangadharan SK. A scoping review of clusters of multiple long-term conditions in people with intellectual disabilities and factors impacting on outcomes for this patient group. *J Intellect Disabil*. (2023) 27:1045–61. doi: 10.1177/17446295221107275
- Yang H, Pawitan Y, Fang F, Czene K, Ye W. Biomarkers and disease trajectories influencing women's health: results from the UK Biobank cohort. *Phenomics*. (2022) 2:184–93. doi: 10.1007/s43657-022-00054-1
- Tyrer F, Dunkley A, Singh J, Kristunas C, Khunti K, Bhaumik S, et al. Multimorbidity and lifestyle factors among adults with intellectual disabilities: a cross-sectional analysis of a UK cohort. *J Intellect Disabil Res*. (2019) 63:255–65. doi: 10.1111/jir.12571
- Van Timmeren E, Waning A, Van Schrojenstein Lantman-de H, Van der Putten A, Van der Schans C. Patterns of multimorbidity in people with severe or profound intellectual and motor disabilities. *Res Dev Disabil*. (2017) 67:28–33. doi: 10.1016/j.ridd.2017.05.002
- De Bhailis ÁM, Kalra PA. Hypertension and the kidneys. *Br J Hosp Med*. (2022) 83:1–11. doi: 10.12968/hmed.2021.0440
- Masenga SK, Kirabo A. Hypertensive heart disease: risk factors, complications and mechanisms. *Front Cardiovasc Med*. (2023) 10:1205475. doi: 10.3389/fcvm.2023.1205475
- Tyrer F, Morris R, Kiani R, Gangadharan SK, Kundaje H, Rutherford MJ. Health needs and their relationship with life expectancy in people with and without intellectual disabilities in England. *Int J Environ Res Public Health*. (2022) 19:6602. doi: 10.3390/ijerph19116602
- Siddiqui N, Dwyer M, Stankovich J, Peterson G, Greenfield D, Si L, et al. Hospital length of stay variation and comorbidity of mental illness: a retrospective study of five common chronic medical conditions. *BMC Health Serv Res*. (2018) 18:1–10. doi: 10.1186/s12913-018-3316-2
- Beck MK, Jensen AB, Nielsen AB, Perner A, Moseley PL, Brunak S. Diagnosis trajectories of prior multi-morbidity predict sepsis mortality. *Sci Rep*. (2016) 6:36624. doi: 10.1038/srep36624

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fdgth.2025.1528882/full#supplementary-material>