European Journal of Public Health, 1–9

https://doi.org/10.1093/eurpub/ckad179

Understanding the conditions included in data-driven patterns of multimorbidity: a scoping review

Luxsena Sukumaran [1,2], Alan Winston [3], Caroline A. Sabin[1,2]

1 Institute for Global Health, University College London, London, UK
2 National Institute for Health Research (NIHR) Health Protection Research Unit (HPRU) in Blood-borne and Sexually Transmitted Infections at University College London, London, UK
3 Department of Infectious Disease, Imperial College London, London, UK

Correspondence: Luxsena Sukumaran, Institute for Global Health, University College London, Royal Free Campus, Rowland Hill Street, London NW3 2PF, UK, e-mail: luxsena.sukumaran.19@ucl.ac.uk

Background: Despite the growing utilization of data-driven methods to investigate multimorbidity patterns, there is currently no consensus or guidance on the conditions to include when identifying patterns. This scoping review aims to systematically examine the nature of conditions included in existing studies using data-driven techniques. **Methods:** A comprehensive search of three electronic databases (MEDLINE, Web of Science and Scopus) was conducted to identify relevant publications from inception to 28 February 2022 using predefined search terms and inclusion/exclusion criteria. The reference lists and citations of relevant papers were also searched. **Results:** Among 7326 search results, 5444 relevant articles were identified. After screening against the eligibility criteria, 60 articles were included in the review. Half of the reviewed studies reported selection criteria for conditions, with prevalence in the population of interest being the most common criterion (40%). Most studies included at least one neurological [59 (98.3%)], musculoskeletal [58 (96.7%)], respiratory [57 (95.0%)] or mental health [56 (93.3%)] condition. In contrast, only a small proportion of studies included skin [17 (28.3%)], infections [14 (23.3%)] or autoimmune conditions [10 (16.7%)]. Nine conditions (hypertension, diabetes, cancer, arthritis, COPD, asthma, depression, stroke and osteoporosis) were included by more than half of the studies. **Conclusions:** This review highlights the considerable heterogeneity among the conditions included in analyses of multimorbidity patterns. Researchers should provide a clear rationale for the selection of conditions to facilitate comparisons across studies and ensure reproducibility, as well as consider selecting a diverse range of conditions to capture the complexity of multimorbidity.

Introduction

Multimorbidity, defined as the presence of multiple health conditions in an individual,[1,2] represents a significant health challenge for individuals and healthcare systems. Medical advances and increases in population ageing have contributed to a rising prevalence of multimorbidity,[3] the health implications of which include poorer quality of life, functional disability and mortality.[4–7] Healthcare systems are largely organized around a single disease-based paradigm and are rarely structured to address the complex healthcare needs associated with multimorbidity, resulting in uncoordinated and fragmented care.[8]

Most multimorbidity studies use simple or weighted disease counts. However, these methods do not account for the large list of possible combinations of health conditions or the fact that some conditions may co-exist due to shared pathophysiological mechanisms and/or risk factors.[9,10] Consequently, data-driven approaches have been increasingly applied to explore the underlying structure in the distribution of co-occurring conditions,[10,11] with cluster analysis (CA) and exploratory factor analysis (EFA) being most commonly used. Multimorbidity patterns can provide an insight into the synergies and effects of specific combinations of conditions, thus informing the development and delivery of targeted interventions/guidelines for improved health outcomes. Previous systematic reviews have found that clinically relevant and replicable patterns can be identified, with patterns of 'cardiovascular and metabolic diseases', 'mental health' and 'musculoskeletal disorders' consistently reported.[10–12] Nonetheless, the frequency of specific combinations/patterns of conditions will be dependent on the list of conditions used. Furthermore, the extent to which the selected conditions accurately capture multimorbidity burden in the population remains unclear. Currently, there is no agreement on the conditions to be used to define multimorbidity. Calls for multimorbidity research to be more person-centered[13,14] necessitate a shift from a biomedical-centred focus of 'what is the matter with the patient' to a patient-centred approach of 'what matters to the patient'. As such, the definition of multimorbidity should consider patients' priorities and concerns, including their physical and psychosocial burden, alongside chronic diseases. For this reason, several studies have included symptoms in their multimorbidity definition as these have been associated with patient-reported health outcomes.[13,15]

This scoping review aims to systematically examine the nature and variation of conditions included in studies on multimorbidity patterns, as well as identify limitations and evidence-gaps for future research that would improve our understanding and identification of data-driven patterns.

Methods

We conducted a scoping review in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-analysis Extension for Scoping Reviews (PRISMA-ScR) protocol.[16]

Search strategy

Relevant studies were identified through the MEDLINE (Ovid), Web of Science (Clarivate Analytics) and Scopus (Elsevier) electronic databases from their inception to 28 February 2022. Search terms

included variations of the term multimorbidity to increase sensitivity: 'multimorbidity', 'multi-morbidity', 'multimorbidities', 'multiple morbidities', 'multiple chronic diseases' and 'multiple chronic conditions'. We also included variations of the term 'comorbidity' since 'multimorbidity' and 'comorbidity' have previously been used synonymously.[17,18] We searched for data-driven studies using the following terms: 'cluster', 'pattern', 'non-random association', among others. The final search strategy is shown in Supplementary table S1.

Eligibility criteria

Inclusion and exclusion criteria were defined prior to database searches (table 1). We included articles that applied a data-driven approach to identify multimorbidity patterns within a given population, with an explicit statement of the conditions considered. We included all studies in the analysis of patterns, regardless of the number of conditions included. To examine whether there were variations in the selection of conditions based on the age of the population of interest, we included studies that focused on adult (aged ≥15 years) multimorbidity. Included studies were written in English language and were published in peer-reviewed journals. We excluded: qualitative and non-original research articles (reviews, meta-analyses, editorials, commentaries or conference presentations); articles that examined patterns without a data-driven approach (e.g. those that reported only the observed–expected ratio or a simple count of conditions); and articles that selected individuals based on the presence of an index condition (i.e. studies of a comorbidity). Our search had no restrictions on data of publication.

Study selection and data extraction

After removal of duplicates, articles were screened for eligibility by title, abstract and full text using Endnote (version 20). The reference lists of the selected articles were manually screened to identify further articles that may have been missed in the main search. The following data items were extracted from eligible studies following full-text review: study characteristics (title, first author, publication date, study design/setting, data source, study population and sample size); definition of multimorbidity (the list of conditions and selection criteria used); and the analytical approach used to identify patterns. The conditions used by studies were compiled and classified by category (diseases, risk factors and symptoms, as used by Willadsen and colleagues[19]) and respective body system (due to inconsistencies in labelling of conditions as some studies use broad system groups instead of specific conditions) for descriptive analyses. 'Diseases' were defined as conditions with an objective diagnosis and/or distinct diagnostic code; 'risk factors' were considered as conditions or measurements associated with the probability of disease or mortality; and 'symptoms' were defined using the International Classification of Primary Care definition. Finally, we extracted the key findings and limitations of each study based on generalizability, methodology and interpretation. To develop the narrative synthesis, we analyzed and summarized the conditions considered when identifying the patterns, investigated similarities and differences between studies and examined the strengths and weaknesses of the approaches. The references of all 60 selected articles are provided in Supplementary table S2.

Results

Figure 1 depicts the study selection process. Of the initial 7326 records identified, 1882 were duplicates and removed. Of the remaining 5444 records, 107 were of potential relevance. After full-text review, 56 records were excluded. The reference lists of the remaining 51 records identified an additional 9 articles for inclusion. In total, 60 papers were selected for this review.

Study characteristics

Key characteristics of the 60 studies are summarized in Supplementary table S3. All studies were published since 2003, with 25 (42%) published in the last 2 years. The sample size varied widely, ranging from 247 to 3 349 721 participants [median 10 759, interquartile range (IQR) 2526–77 782]. Half of the included studies were primarily based on European data. The majority of studies had age restrictions, with 73% ($n = 44$) of studies focusing on older populations (aged ≥50 years), although three articles included participants as young as 15 years. Data on health conditions were mainly collected through survey/questionnaires [interview-based ($n = 31$); self-report ($n = 9$)], followed by electronic health records (EHR; $n = 12$) and administrative sources ($n = 4$). In four studies, survey data were combined with EHR, administrative sources or clinical examination data.

Analytical methods

The most common data-driven approach employed to identify multimorbidity patterns was latent class analysis (LCA) ($n = 22$), followed by EFA ($n = 19$) and CA ($n = 16$) (Supplementary table S3).

Table 1 Study inclusion/exclusion criteria

Criteria	Inclusion	Exclusion
Study design	Quantitative studies (e.g. prospective and retrospective cohort studies) Cross-sectional or longitudinal studies	Non-original research articles including: i. Systematic reviews/meta-analyses ii. Qualitative studies iii. Editorials iv. Commentaries v. Conference presentations
Methodology	Measure of multimorbidity patterns using a statistical technique, e.g. i. Cluster analysis ii. Exploratory factor analysis iii. Latent class analysis iv. Principal component analysis Conditions included in analysis explicitly stated	Measure of multimorbidity patterns using simple counts, weighted indices or observed–expected ratios Conditions not listed Analysis based on presence of an index condition (studies of comorbidity)
Study population	Adults aged ≥15 years	Infants, children and/or adolescents (<15 years) Animal research
Publication type	Peer-reviewed journal articles Written in English	Grey literature Not written in English

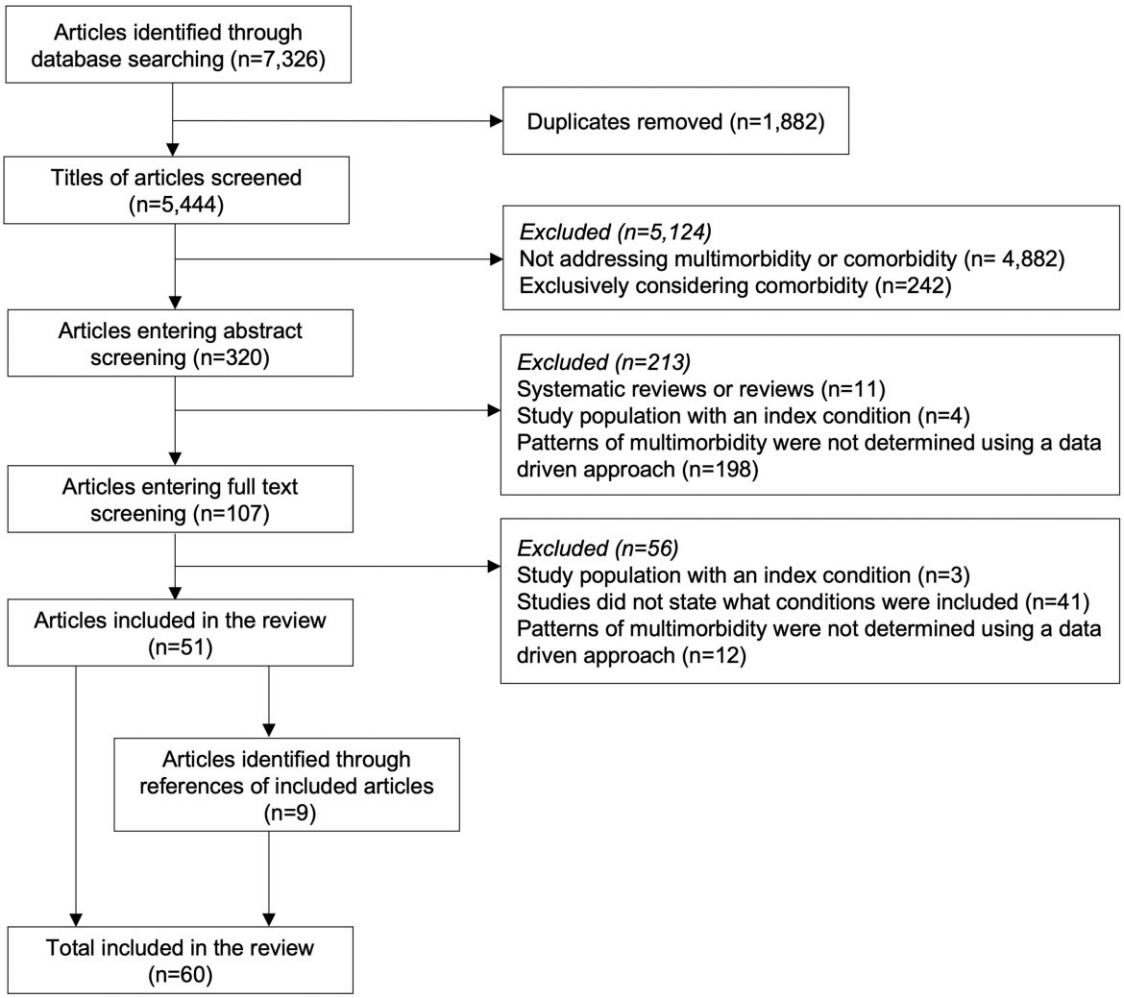

Figure 1 Study selection for scoping review

Two studies used multiple statistical techniques: (i) CA, PFA and LCA, and (ii) LCA and EFA. Twenty studies (33%) stratified their patterns by age, sex, race and/or region: 5/22 LCA; 9/19 EFA; 5/16 CA; and 1/1 LCA and EFA (Supplementary table S4). Notably, among the studies that included younger populations ($n = 15$), only 33% stratified their patterns by age.

Multimorbidity patterns

The number of patterns identified ranged from 2 to 8, with an overall median of four (IQR 3–5). LAC studies identified from 3 to 8 patterns (median 4, IQR 3–5.3), EFA studies 2–6 (median 3, IQR 3–4) and CA studies 3–8 (median 5, IQR 4–6), with the two PFA studies both identifying three patterns. Generally, there was reasonable comparability in the patterns identified among the CA and EFA studies (Supplementary figure S5). Three patterns were consistently observed: 'cardiovascular and/or metabolic', 'mental health' and 'musculoskeletal'. 'Cardio-metabolic' and 'musculoskeletal' patterns were also identified across LFA studies, alongside a comparative 'healthy/low risk' pattern and a pattern including at least one respiratory condition. The three patterns identified in the studies using PFA included both a 'cardiometabolic' and 'mental health' pattern, and either a 'geriatric' or a 'chronic disease' pattern.

List of conditions included

Selection criteria for their list of conditions were provided in half of the studies reviewed. Twenty-one studies used prevalence as a criterion for selecting conditions, with the majority (81%) utilizing a threshold cut-off (ranging from 1% to 5%). Four studies used definitions from previous publications to develop their list of conditions. Two studies specifically selected conditions associated with health outcomes, such as quality of life, everyday functioning and high economic costs. Furthermore, three studies used a combination of both prevalence and guidance from previous publications as criteria for selecting their list of conditions.

Over half of the studies ($n = 39$, 63%) used *ad hoc* lists for ascertainment of conditions. The number of conditions included ranged from 8 to 114 (median 16, IQR 12–27) (Supplementary table S2). Four studies employed the same list of 114 chronic expanded diagnostic clusters that was published by Salisbury et al.[19] in 2011. The median (IQR) number of conditions selected varied across the different data sources utilized: survey/questionnaire [14 (11–22)], EHR [40 (24–57)], administrative data [34 (22–63)] and a combination of two data sources [16 (14–18)]. Furthermore, the median (IQR) number of conditions also varied depending on the statistical technique employed: CA [18 (15–60)], EFA [21 (12–41)], LC [13 (11–16)] and PFA [14 (17–19)].

The categories of conditions used also varied across the reviewed studies. In almost all (98%) of the reviewed studies, at least one risk factor was included, with hypertension (98%), osteoporosis (57%) and obesity (40%) being the most commonly reported. In addition, over half (56%) of these studies also included at least one symptom, with chronic pain (35%) and migraines/headaches (22%) being the most frequently used. All 60 studies included at least one cardiovascular and metabolic/endocrine condition (figure 2). Most studies also included at least one condition from the following categories in their

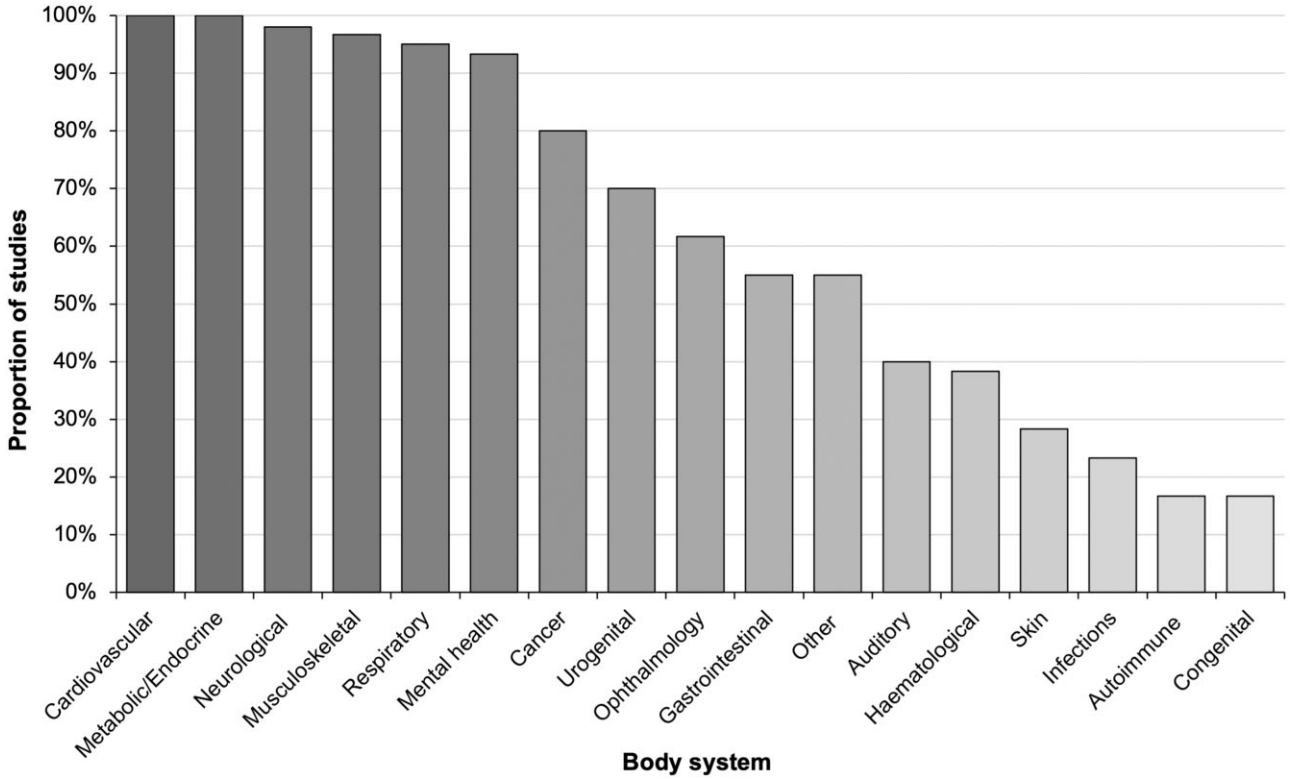

Figure 2 Conditions grouped by body system included in the reviewed studies ($n = 60$)

lists: neurological [59 (98.3%)], musculoskeletal [58 (96.7%)], respiratory [57 (95.0%)], mental health [56 (93.3%)] and cancer [48 (80.0%)]. In contrast, only a small proportion of studies included: skin [17 (28.3%)], infections [14 (23.3%)], autoimmune [10 (16.7%)] and congenital [10 (16.7%)] conditions. In terms of individual conditions, nine conditions were included in the lists used by more than half of the studies (table 2). Hypertension and diabetes were both included by 59 (98.3%) studies, with cancer [50 (83.3%)], arthritis [47 (78.3%)], COPD/bronchitis/emphysema [42 (72.0%)], asthma [40 (66.7%)], depression [39 (65.0%)], stroke [38 (63.3%)] and osteoporosis [34 (56.7%)] also being commonly included. In contrast, health conditions that were selected by <5% of studies included: insomnia [2 (3.3%)], dizziness [2 (3.3%)], tinnitus [2 (3.3%)], autism [1 (1.7%)], sciatica [1 (1.7%)] and fatigue [1 (1.7%)], among others. Most studies [56 (93.3%)] included both physical and mental health conditions, while four (6.7%) studies focused solely on physical conditions. Over half of the studies [56 (93.3%)] included at least one mental health condition, with the most common conditions being depression [39 (65.0%)], anxiety [15 (25.0%)] and schizophrenia [14 (23.3%)]. Other conditions, including bipolar disorder, alcohol-related disorders, panic attacks, eating disorders and obsessive-compulsive disorder, were incorporated in <5% of studies. Furthermore, all studies that included younger populations ($n = 15$) considered at least one comorbidity typically associated with this demographic. Notably, asthma, anxiety and/or depression, obesity and diabetes were most commonly selected. In contrast, other potentially relevant conditions, such as skin disorders (e.g. psoriasis/eczema) and sexually transmitted diseases, were either less common or excluded in these studies.

Discussion

Data-driven approaches are becoming more frequently employed to study the underlying distribution between co-occurring conditions. This scoping review identified 60 articles studying multimorbidity patterns and found significant variation in the type of conditions included. These studies have also highlighted several challenges for further discussion, including clarity around reporting of the criteria used to select conditions and around the framework used to define multimorbidity.

Half of the reviewed studies provided a selection criterion or criteria for their list of conditions. The most common criterion used was highly prevalent conditions in either the general or study population. Although the majority of these studies used a prevalence cut-off threshold, they did not provide a justification for their cut-off choice. There is no consensus on what constitutes as an appropriate cut-off threshold, and different cut-offs may lead to the identification of different patterns. Therefore, researchers should provide a clear rationale for their cut-off choice, as well as conduct sensitivity analyses with different cut-offs to explore the robustness of their findings. While the selection of highly prevalent conditions is based on the assumption that these conditions are more likely to co-occur, this approach may not necessarily capture the conditions with the most significant implications for patient-important health outcomes. Thus, the use of multiple selection criteria may be more appropriate to identify patterns that can effectively guide clinical decision-making. The use of *ad hoc* lists in the remaining studies may be attributed to the arbitrary and often criticized nature of defining selection criteria. These studies may have also been limited pragmatically by what was available from their data source. Nevertheless, researchers should be transparent about these issues to improve the transparency and rigour of research on multimorbidity patterns.

There was large variation in the conditions included in the assessment of multimorbidity patterns. The majority of studies included at least one risk factor (98%) and/or symptom (57%) in their list of conditions. Hypertension was the most frequent risk factor and was notably included in more studies than diseases, such as depression and cancer. Although risk factors may be common in a given population and help identify individuals who may develop future illnesses, their inclusion may lead to 'double counting' for diseases that are already present and may have little impact on current morbidity burden. Symptoms, on the other hand, can have a significant impact

Table 2 Individual conditions included in studies of multimorbidity patterns

Organ System	Condition	Category (D/R/S)	Number of studies	Frequency (%)
Cardiovascular	Hypertension	R	59	98.3
	Stroke	D	38	63.3
	Coronary heart disease	D	27	45
	Congestive heart failure	D	24	40
	Arrhythmia	D	18	30
	Dyslipidaemia	R	18	30
	Myocardial infarction	D	14	23.3
	Atrial fibrillation	D	12	20
	Cardiac valve diseases	D	11	18.3
	Angina	D	9	15
	Atherosclerosis	D	8	13.3
	Hypercholesterolaemia	R	7	11.7
	Cardiomyopathies	D	5	8.3
	Aortic aneurysm	D	4	6.7
	Cardiac insufficiency	D	1	1.7
	Heart murmurs	S	1	1.7
	Hypotension	R	1	1.7
	Peripheral arterial disease	D	1	1.7
	Pulmonary heart disease	D	1	1.7
Metabolic/endocrine	Diabetes	D	59	98.3
	Thyroid disease	D	27	45
	Obesity	R	24	40
	Lipoprotein metabolism disorders	D	10	16.7
	Hypertriglyceridaemia	R	2	3.3
	Hypoalphalipoproteinaemia	R	1	1.7
Respiratory	COPD/bronchitis/emphysema	D	42	70
	Asthma	D	40	66.7
	Allergies	D	13	21.7
	Pulmonary embolism	D	4	6.7
	Tracheostomy	D	4	6.7
	Bronchiectasis	D	2	3.3
	Chronic sinusitis	D	2	3.3
	Breathing difficulties	S	1	1.7
	Wheezing	S	1	1.7
Musculoskeletal	Arthritis	D	47	78.3
	Osteoporosis	R	34	56.7
	Spondylopathies/dorsopathies	D	11	18.3
	Injuries/work-related disorders	D	9	15
	Arthropathy	D	7	11.7
	Gout	D	6	10
	Fractures	D	4	6.7
	Muscular dystrophy	D	4	6.7
	Quadriplegia and paraplegia	D	4	6.7
	Kyphoscoliosis	D	4	6.7
	Back/neck problems	S	1	1.7
Mental health	Depression	D	39	65
	Anxiety	D	15	25
	Schizophrenia	D	14	23.3
	Sleep problems	S	10	16.7
	Substance use-related diseases	D	8	13.3
	Alcohol-related diseases	D	5	8.3
	Bipolar disorder	D	2	3.3
	Eating disorders	D	2	3.3
	Insomnia	D	2	3.3
	Obsessive-compulsive disorder	D	2	3.3
Cancer	Cancers	D	50	83.3
Neurological	Dementia	D	24	40
	Parkinson's	D	21	35
	Cerebrovascular disease	D	15	25
	Migraine/headaches	S	13	21.7
	Multiple sclerosis	D	12	20
	Epilepsy	D	11	18.3
	Peripheral neuropathy	D	11	18.3
	Attention deficit disorder	D	4	6.7
	Cerebral palsy	D	4	6.7
	Memory impairments	D	4	6.7
	Paralytic syndromes	D	4	6.7
	Seizure disorder	D	4	6.7
	Developmental disorder	D	4	6.7
	Alzheimer's disease	D	3	5
	Cognitive impairment	D	3	5
	Brain infarction/haemorrhage	D	2	3.3
	Dizziness	S	2	3.3
	Autism	D	1	1.7

(continued)

Table 2 Continued

Organ System	Condition	Category (D/R/S)	Number of studies	Frequency (%)
	Fibromyalgia	D	1	1.7
	Learning disability	D	1	1.7
	Sciatica	D	1	1.7
Gastrointestinal	Chronic liver disease	D	17	28.3
	Inflammatory bowel disease	D	13	21.7
	Gastroesophageal reflux disease	D	6	10
	Irritable bowel syndrome	D	6	10
	Colitis	D	5	8.3
	Diverticular disease	D	5	8.3
	Hepatitis	D	5	8.3
	Chronic pancreatitis	D	4	6.7
	Lactose intolerance	D	4	6.7
	Gastritis	D	3	5
	Constipation	S	2	3.3
	Heartburn	S	2	3.3
	Stomach ulcer	D	2	3.3
	Chronic cholecystitis/gallstones	D	1	1.7
	Adiposis	D	1	1.7
	Crohn's	D	1	1.7
	Dyspepsia	D	1	1.7
Urogenital	Chronic kidney disease (CKD)	D	21	35
	Prostatic hypertrophy	D	8	13.3
	Chronic renal failure	D	7	11.7
	Urinary incontinence	S	6	10
	Renal calculi	D	5	8.3
	Endometriosis	D	4	6.7
	Nephritis	D	4	6.7
	Prostatitis	D	4	6.7
	Utero-vaginal prolapse	D	4	6.7
	Vesicoureteral reflux	D	4	6.7
	Renal insufficiency	D	2	3.3
	Haemorrhoids	D	1	1.7
	Sexual dysfunction	D	1	1.7
	Urinary tract calculi	D	1	1.7
Ophthalmological	Cataracts	D	20	33.3
	Glaucoma	D	15	25
	Blindness	D	11	18.3
	Vision loss/impairment	S	9	15
	Diabetic retinopathy	D	5	8.3
Infections	HIV/AIDS	D	6	10
	Tuberculosis	D	7	11.7
	Meningitis	D	1	1.7
Haematological	Anaemia	D	21	35
	Peripheral vascular disease	D	11	18.3
	Varicose veins	D	6	10
	Iron deficiencies	D	6	10
	Deep vein thrombosis	D	5	8.3
	Haematological neoplasms	D	5	8.3
	Venous and lymphatic diseases	D	5	8.3
	Haemophilia	D	4	6.7
	Blood clotting disorder	D	1	1.7
	Phlebitis and thrombophlebitis	D	1	1.7
Skin disorders	Skin ulcers	D	13	21.7
	Psoriasis	D	7	11.7
	Dermatitis/eczema	D	6	10
	Fibrocystic breast disease	D	4	6.7
	Disorders of hair and follicles (e.g. alopecia)	D	4	6.7
Congenital	Congenital anomalies of limbs, hands and feet	D	5	8.3
	Cleft lip and palate	D	4	6.7
	Congenital heart disease	D	4	6.7
	Congenital hip dislocation	D	4	6.7
	Hypospadias	D	4	6.7
Other	Deafness, hearing loss/impairment	S	24	40
	Chronic pain	S	21	35
	Autoimmune diseases	D	15	25
	Chromosomal abnormalities	D	9	15
	Solid neoplasms	D	5	8.3
	Teeth/gums-related	D	4	6.7
	Cystic fibrosis	D	4	6.7
	Transplant status	R	4	6.7
	Surgical aftercare	R	4	6.7
	Disabilities	D	2	3.3
	Tinnitus	D	2	3.3
	Clumsiness	S	1	1.7
	Fatigue	S	1	1.7

Note: D, disease; R, risk factor; S, symptom.

on patients' quality of life and function, even in the absence of a diagnosed disease, and thus their inclusion may provide a more patient-centred understanding of multimorbidity patterns. In addition, a number of studies were inconsistent on whether conditions were included as a broader heterogenous group (e.g. cardiovascular disease) or included as individual entities (e.g. angina, arrhythmia and cardiac insufficiency). Similarly, some studies included cancer as one disease, while other studies distinguished between cancers (i.e. skin vs. other cancers) and/or different forms of malignant neoplasms and included them as separate diseases. Almost all studies (95%) included at least one cardiovascular, endocrine, musculoskeletal and respiratory condition, whereas skin-related, infections, autoimmune and birth defects were included by <30% studies. Nine conditions were included by more than half of the studies, eight of which were physical conditions (hypertension, diabetes, cancer, arthritis, COPD, asthma, depression, stroke and osteoporosis). Four studies did not include any mental health conditions, which suggests improvements in recognition that mental health plays a key role in multimorbidity. However, only depression ($n = 39$) and anxiety ($n = 25$) were consistently included in studies, while other mental health conditions (e.g. sleeping problems, alcohol-related, bipolar, eating disorders, etc.), recognized to also affect health and quality of life, were included by less than a third of studies. This pattern was also observed among other groups of conditions. For example, studies that included nervous system disorders predominantly included Parkinson's and dementia, while conditions, such as epilepsy, ADHD and peripheral neuropathy, were included in 5–18% of studies. This reflects that studies primarily focus on conditions that are commonly reported in their data source/literature, and/or those that are highly prevalent within the population of interest. However, this approach may lead to the oversight of conditions with a stronger association with adverse and/or patient-important health outcomes.

Strengths and weaknesses of this review

To our knowledge, this study is the first to review the nature of conditions included in studies using a data-driven approach to identify multimorbidity patterns. Our review identified a larger database of articles examining data-driven patterns (60 studies compared with 6 and 14 articles identified by Violan et al.[11] and Prados-Torres et al.,[10] respectively). This can be attributed to the comprehensive database search we employed, which incorporated variations of the terms 'comorbidity' and 'multimorbidity' and allowed us to identify additional papers that were not detected by earlier reviews. In addition, previous reviews have primarily focused on patterns in either primary care settings or limited their search to studies that selected ≥10 conditions.[10,11] However, our search strategy did not impose such restrictions to maximize identification of studies and limit selection bias. Nevertheless, our study also has limitations. The classification of the conditions (diseases, risk factors and symptoms) in this review can be debated. However, we were limited in certain cases due to the variations in how individual studies coded their conditions. For instance, some studies had grouped migraines and headaches together, while others may argue that they should be treated separately, with migraine being classified as a disease and headaches as a symptom.[20] Additionally, all the reviewed studies were based in high- or middle-income countries, and therefore multimorbidity patterns in low-income countries could not be evaluated. These countries face a greater burden to communicable health conditions, including chronic infectious diseases (e.g. TB and HIV/AIDS), which we recognize will play a larger role in multimorbidity patterns, compared with that of high-income countries. Furthermore, we excluded studies that measured multimorbidity using simple counts or weighted indices, which may be considered a limitation as these methods are more commonly used in existing literature and clinical practice. However, these methods cannot separate coincident (random) comorbidity from non-random comorbidity and understanding the latter will allow us to study interactions and shared aetiologies between comorbidities.

Comparison with previous literature

Consistent with our review, Prados-Torres et al.[10] previously reported that COPD, hypertension and diabetes were more frequently selected by studies looking at data-driven patterns. However, our review provides a more comprehensive updated search, a decade later, identifying a larger number of relevant studies (60 vs. 14). Our review also provides a more detailed insight into the implications of conditions selected in analyses of multimorbidity patterns. Multimorbidity frameworks have been proposed by previous systematic reviews to guide studies using simple or weighted disease counts, but their relevance for studies utilizing data-driven approaches has not been explored. For example, Diederichs et al.[18] proposed a framework comprising at least 11 conditions (both diseases and risk factors), whereas both Fortin et al.[21] and Willadsen et al.[19] highlighted the importance of including conditions associated with patient's health outcomes, such as symptoms, to capture the clinical reality of individuals living with multimorbidity. In our review, we found that 57% of studies included symptoms in their analyses of multimorbidity patterns, suggesting a growing recognition of their importance to capture the complexity of multimorbidity.

Implications for clinical practice, health policy and future research

Multimorbidity is a growing public health concern, and it is important that studies continue to study data-driven patterns. By doing so, we can identify patterns associated with a greater burden and deliver holistic and targeted care to individuals exhibiting these patterns. Our key recommendation is that researchers should be explicit about the conditions they select and provide a rationale for their inclusion. Researchers should also consider conditions that may be of low prevalence and/or those typically under-reported. Identifying these conditions may be problematic within routine and self-reported data, as often self-reported interviews and questionnaires are simplified to be comprehensible from a lay perspective. Similarly, health records and registers are often generalized and focus on frequently registered health conditions, conditioned by their prevalence and severity (e.g. cancers and cardiovascular diseases), which prevents rare/complex conditions, such as autoimmune and mental health disorders, to be detected. However, recognizing that these data sources may reduce our chances of identifying those at greatest burden (e.g. higher healthcare utilization, functional decline and poor quality of life) is important. To overcome this issue, researchers should consider the following: (i) combining data sources, e.g. using survey data with administrative sources and/or clinical examinations; (ii) identifying patterns using different prevalence cut-off thresholds; and (iii) using multiple selection criteria, e.g. highly prevalent conditions and conditions associated with patient-important outcomes (based on previous literature). Studies should also recognize that selecting the most prevalent conditions may not be appropriate to accurately assess multimorbidity patterns across different population subgroups. For example, many of these prevalent conditions (such as cardiovascular and musculoskeletal disorders) are more frequently observed among older populations. However, multimorbidity is becoming increasingly prevalent among younger populations, who are likely to present different multimorbidity patterns that may include under-reported conditions, such as mental health and sexually transmitted diseases. Furthermore, researchers examining both populations should explore and report age-stratified patterns.

Conclusion

In conclusion, this review highlights the wide variation in the nature of health conditions considered in multimorbidity studies, with only

half providing a rationale for their selection. There is a need for greater transparency and detail when reporting the list of conditions and the rationale used, to advance towards a more uniform methodology. While some studies have focused on a limited set of chronic conditions, a holistic approach that incorporates a broader range of health conditions may help us transition from a disease- to person-centred definition of multimorbidity, supporting more accurate identification of individuals at greatest risk for the adverse outcomes associated with multimorbidity.

Supplementary data

Supplementary data are available at *EURPUB* online.

Author contributions

L.S. performed the literature search; screened the titles, abstracts and full texts; independently extracted all the data; and wrote the first draft of the review. C.A.S. and A.W. developed the idea for the review with L.S. and supervised the research process. All authors critically revised, edited and approved the final draft of the manuscript. All authors had final responsibility for the decision to submit for publication. The corresponding author confirms that all listed authors meet authorship criteria.

Acknowledgements

We acknowledge members of the NIHR HPRU in Blood Borne and Sexually Transmitted Infections Steering Committee: Professor Caroline Sabin (HPRU Director), Dr John Saunders (UKHSA Lead), Professor Catherine Mercer, Dr Hamish Mohammed, Professor Greta Rait, Dr Ruth Simmons, Professor William Rosenberg, Dr Tamyo Mbisa, Professor Rosalind Raine, Dr Sema Mandal, Dr Rosamund Yu, Dr Samreen Ijaz, Dr Fabiana Lorencatto, Dr Rachel Hunter, Dr Kirsty Foster and Dr Mamoona Tahir.

Funding

L.S. was funded through the National Institute for Health and Care Research Health Protection Research Unit (NIHR HPRU) in Blood Borne and Sexually Transmitted Infections at University College London in partnership with the UK Health Security Agency (HPRU Grant No: NIHR200911).

Disclaimer

The views expressed are those of the authors and not necessarily those of the NIHR, the Department of Health and Social Care or UKHSA.

Conflicts of interest: C.A.S. reports receipt of funding from Gilead Sciences and ViiV Healthcare for membership of Advisory Boards and for preparation of educational materials. A.W. has received speaker fees, advisory board honoraria or grants via Imperial College London from Gilead Sciences, ViiV Healthcare, MSD and Janssen. L.S. reports no conflicts of interest.

Data availability

The authors confirm that the data supporting the findings of this study are available within the article and/or its Supplementary material.

Key points

- This scoping review highlights the wide variation in the nature of health conditions (diseases, risk factors and symptoms) considered in studies analyzing multimorbidity patterns.
- To capture the true burden of multimorbidity, we recommend researchers apply a patient-centred approach when selecting conditions for analyses of multimorbidity patterns, i.e. consider conditions that have significant implications for patient-important outcomes.
- Where possible, researchers should use multiple selection criteria and data sources to ensure that under-reported/complex conditions are not overlooked.

References

1 van den Akker M, Buntinx F, Knottnerus JA. Comorbidity or multimorbidity. *Eur J Gen Pract* 1996;2:65–70.

2 The Academy of Medical Sciences. Multimorbidity: a priority for global health research. London: Sciences TAoM ed; 2018. p. 1–142.

3 World Health Organisation. Strategy and action plan for healthy ageing in Europe, 2012–2020. Copenhagen: World Health Organisation Regional Office for Europe, 2012.

4 Ryan A, Wallace E, O'Hara P, Smith SM. Multimorbidity and functional decline in community-dwelling adults: a systematic review. *Health Qual Life Outcomes* 2015;13:168.

5 Fortin M, Lapointe L, Hudon C, et al. Multimorbidity and quality of life in primary care: a systematic review. *Health Qual Life Outcomes* 2004;2:51.

6 Fabbri E, Zoli M, Gonzalez-Freire M, et al. Aging and multimorbidity: new tasks, priorities, and frontiers for integrated gerontological and clinical research. *J Am Med Dir Assoc* 2015;16:640–7.

7 Villacampa-Fernández P, Navarro-Pardo E, Tarín JJ, Cano A. Frailty and multimorbidity: two related yet different concepts. *Maturitas* 2017;95:31–5.

8 Moffat K, Mercer SW. Challenges of managing people with multimorbidity in today's healthcare systems. *BMC Fam Pract* 2015;16:129.

9 Valderas JM, Starfield B, Sibbald B, et al. Defining comorbidity: implications for understanding health and health services. *Ann Fam Med* 2009;7:357–63.

10 Prados-Torres A, Calderon-Larranaga A, Hancco-Saavedra J, et al. Multimorbidity patterns: a systematic review. *J Clin Epidemiol* 2014;67:254–66.

11 Violan C, Foguet-Boreu Q, Flores-Mateo G, et al. Prevalence, determinants and patterns of multimorbidity in primary care: a systematic review of observational studies. *PLoS One* 2014;9:e102149.

12 Ng SK, Tawiah R, Sawyer M, Scuffham P. Patterns of multimorbid health conditions: a systematic review of analytical methods and comparison analysis. *Int J Epidemiol* 2018;47:1687–704.

13 Griffith LE, Gilsing A, Mangin D, et al. Multimorbidity frameworks impact prevalence and relationships with patient-important outcomes. *J Am Geriatr Soc* 2019;67:1632–40.

14 Iris SSH, Amaya A-L, Ashley A, et al. Measuring multimorbidity in research: Delphi consensus study. *BMJ Med* 2022;1:e000247.

15 Nützel A, Dahlhaus A, Fuchs A, et al. Self-rated health in multimorbid older general practice patients: a cross-sectional study in Germany. *BMC Fam Pract* 2014;15:1.

16 Tricco AC, Lillie E, Zarin W, et al. PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. *Ann Intern Med* 2018;169:467–73.

17 Nicholson K, Makovski TT, Griffith LE, et al. Multimorbidity and comorbidity revisited: refining the concepts for international health research. *J Clin Epidemiol* 2019;105:142–6.

18 Diederichs C, Berger K, Bartels DB. The measurement of multiple chronic diseases-a systematic review on existing multimorbidity indices. *J Gerontol A Biol Sci Med Sci* 2011;66:301–11.

19 Willadsen TG, Bebe A, Koster-Rasmussen, R et al. The role of diseases, risk factors and symptoms in the definition of multimorbidity - a systematic review. *Scand J Prim Health Care* 2016;34:112–21.

20 Eigenbrodt AK, Ashina H, Khan S, et al. Diagnosis and management of migraine in ten steps. *Nat Rev Neurol* 2021;17:501–14.

21 Fortin M, Almirall J, Nicholson K. Development of a research tool to document self-reported chronic conditions in primary care. *J Comorb* 2017;7: 117–23.