Identifying and visualising multimorbidity and comorbidity patterns in patients in the English National Health Service: a population-based study

*Valerie Kuan, Spiros Denaxas, Praveetha Patalay, Dorothea Nitsch, Rohini Mathur, Arturo Gonzalez-Izquierdo, Reecha Sofat, Linda Partridge, Amanda Roberts, Ian C K Wong, Melanie Hingorani, Nishi Chaturvedi, Harry Hemingway, Aroon D Hingorani, on behalf of the Multimorbidity Mechanism and Therapeutic Research Collaborative (MMTRC)**

Summary

Background Globally, there is a paucity of multimorbidity and comorbidity data, especially for minority ethnic groups and younger people. We estimated the frequency of common disease combinations and identified non-random disease associations for all ages in a multiethnic population.

Methods In this population-based study, we examined multimorbidity and comorbidity patterns stratified by ethnicity or race, sex, and age for 308 health conditions using electronic health records from individuals included on the Clinical Practice Research Datalink linked with the Hospital Episode Statistics admitted patient care dataset in England. We included individuals who were older than 1 year and who had been registered for at least 1 year in a participating general practice during the study period (between April 1, 2010, and March 31, 2015). We identified the most common combinations of conditions and comorbidities for index conditions. We defined comorbidity as the accumulation of additional conditions to an index condition over an individual's lifetime. We used network analysis to identify conditions that co-occurred more often than expected by chance. We developed online interactive tools to explore multimorbidity and comorbidity patterns overall and by subgroup based on ethnicity, sex, and age.

Findings We collected data for 3 872 451 eligible patients, of whom 1 955 700 (50·5%) were women and girls, 1 916 751 (49·5%) were men and boys, 2 666 234 (68·9%) were White, 155 435 (4·0%) were south Asian, and 98 815 (2·6%) were Black. We found that a higher proportion of boys aged 1–9 years (132 506 [47·8%] of 277 158) had two or more diagnosed conditions than did girls in the same age group (106 982 [40·3%] of 265 179), but more women and girls were diagnosed with multimorbidity than were boys aged 10 years and older and men (1 361 232 [80·5%] of 1 690 521 *vs* 1 161 308 [70·8%] of 1 639 593). White individuals (2 097 536 [78·7%] of 2 666 234) were more likely to be diagnosed with two or more conditions than were Black (59 339 [60·1%] of 98 815) or south Asian individuals (93 617 [60·2%] of 155 435). Depression commonly co-occurred with anxiety, migraine, obesity, atopic conditions, deafness, soft-tissue disorders, and gastrointestinal disorders across all subgroups. Heart failure often co-occurred with hypertension, atrial fibrillation, osteoarthritis, stable angina, myocardial infarction, chronic kidney disease, type 2 diabetes, and chronic obstructive pulmonary disease. Spinal fractures were most strongly non-randomly associated with malignancy in Black individuals, but with osteoporosis in White individuals. Hypertension was most strongly associated with kidney disorders in those aged 20–29 years, but with dyslipidaemia, obesity, and type 2 diabetes in individuals aged 40 years and older. Breast cancer was associated with different comorbidities in individuals from different ethnic groups. Asthma was associated with different comorbidities between males and females. Bipolar disorder was associated with different comorbidities in younger age groups compared with older age groups.

Interpretation Our findings and interactive online tools are a resource for: patients and their clinicians, to prevent and detect comorbid conditions; research funders and policy makers, to redesign service provision, training priorities, and guideline development; and biomedical researchers and manufacturers of medicines, to provide leads for research into common or sequential pathways of disease and inform the design of clinical trials.

Funding UK Research and Innovation, Medical Research Council, National Institute for Health and Care Research, Department of Health and Social Care, Wellcome Trust, British Heart Foundation, and The Alan Turing Institute.

Lancet Digit Health 2022

Published Online November 29, 2022 https://doi.org/10.1016/S2589-7500(22)00187-X

*Consortium members are listed in the appendix (pp 94–95)

Institute of Health Informatics (V Kuan PhD, Prof S Denaxas PhD, A Gonzalez-Izquierdo PhD, Prof H Hemingway FMedSci), **Health Data Research UK** (V Kuan, Prof S Denaxas, A Gonzalez-Izquierdo, Prof H Hemingway), **UCL BHF Research Accelerator** (Prof S Denaxas, Prof A D Hingorani PhD), **Centre for Longitudinal Studies** (Prof P Patalay PhD), **MRC Unit for Lifelong Health and Ageing at UCL** (Prof P Patalay, Prof N Chaturvedi MD), **Institute of Healthy Ageing, Department of Genetics, Evolution and Environment** (Prof L Partridge PhD), **School of Pharmacy** (Prof I C K Wong PhD), **and Institute of Cardiovascular Science** (Prof A D Hingorani), **University College London, London, UK; Alan Turing Institute, London, UK** (Prof S Denaxas); **University College London Hospitals NIHR Biomedical Research Centre, London, UK** (Prof S Denaxas, Prof H Hemingway, Prof A D Hingorani); **Department of Non-Communicable Disease Epidemiology, London School of Hygiene & Tropical Medicine, London, UK** (Prof D Nitsch MD, Prof R Mathur PhD); **Centre for Primary Care, Wolfson Institute of Primary Care, Queen Mary University of London, London, UK** (Prof R Mathur); **Department of Pharmacology and Therapeutics, University of Liverpool, Liverpool, UK** (Prof R Sofat PhD); **British Heart Foundation Data Science Centre, HDR UK, London, UK**

Introduction

Multimorbidity, canonically defined as the coexistence of two or more chronic conditions,[1,2] presents a challenge to medical practice and research, with treatment guidelines largely based on clinical trials that routinely exclude individuals with multiple conditions.[3]

Accrual of multimorbidity can be considered from two distinct perspectives: frequency and non-random

(Prof S Denaxas, Prof R Sofat); **Max Planck Institute for Biology of Ageing, Cologne, Germany** (Prof L Partridge); **Nottingham Support Group for Carers of Children with Eczema, Nottingham, UK** (A Roberts BSc); **Centre for Safe Medication Practice and Research, Department of Pharmacology and Pharmacy, LKS Faculty of Medicine, University of Hong Kong, Hong Kong Special Administrative Region, China** (Prof I C K Wong); **Aston Pharmacy School, Aston University, Birmingham, UK** (Prof I C K Wong); **Moorfields Eye Hospital, London, UK** (M Hingorani FRCOphth)

Correspondence to: Dr Valerie Kuan, Institute of Health Informatics, University College London, London NW1 2DA, UK
v.kuan@ucl.ac.uk

See Online for appendix

Research in context

Evidence before this study

We searched MEDLINE on Oct 26, 2022, with no language restrictions, for studies published since database inception describing multimorbidity or comorbidity patterns using the terms "multimorbidity" OR "multi-morbidity" OR "comorbidity" OR "co-morbidity" AND "network" OR "cluster". Our search returned 5372 results. Studies of multimorbidity and comorbidity patterns varied greatly in terms of definition of multimorbidity; number, type, and granularity of component conditions; case definition of component conditions (raw disease codes *vs* curated phenotypes); case ascertainment (surveys or self-report *vs* clinical or administrative records); measurement of multimorbidity (weighted *vs* unweighted counts); statistical methods used to group conditions (eg, cluster algorithms, factor analyses, and network analyses); and sample populations (eg, insurance databases, primary *vs* secondary care, and research cohorts). Many studies were based on unrepresentative study populations with restricted age ranges. Studies that covered a wide range of conditions usually used raw and uncurated disease codes, while those using specific disease terms were often limited to fewer than 100 conditions. Some studies reported on non-random disease associations, while others reported on the prevalence of multimorbidity and the frequency of the most common disease combinations. Few investigated multimorbidity patterns in young people or different ethnic groups—evidence gaps identified in the Academy of Medical Sciences report on multimorbidity.

Added value of this study

We present a dual perspective on multimorbidity by examining the most frequent combinations of health conditions, and non-random associations between health conditions that might reflect shared risk factors for 308 curated case definitions for mental and physical health conditions across all organ systems in a representative, multiethnic population in England, stratified by ethnicity, sex, and age, thus addressing evidence gaps in multimorbidity research for minority ethnic groups and younger patients. We also developed online tools to explore multimorbidity and comorbidity patterns by ethnicity, sex, and age. Our findings and tools are an important step towards meeting the challenge of mapping disease clusters.

Implications of all the available evidence

By identifying common disease combinations, we provide information that might encourage clinical trial investigators to actively recruit patients with common comorbidity profiles so that trial data can become more relevant to patients in the real world and clinical guidelines and health-care services can support and provide evidence-based treatment for patients with multiple conditions.

association. Common combinations of conditions might occur randomly because the individual component conditions are common. Identification of these common combinations could inform the design of clinical trials and guidelines to ensure their relevance to patients with multimorbidity and optimise service configuration. Alternatively, conditions can be associated non-randomly and co-occur more often than expected by chance because of shared genetic or environmental determinants, or because one condition or its treatment causes another. Understanding which conditions are non-randomly associated and why could inform preventive and therapeutic interventions for multiple conditions collectively.

To effectively study the impact of multimorbidity on clinical management and shared mechanisms of disease, we extended the definition of multimorbidity to the ever occurrence of two or more conditions in the same individual. This approach is of practical importance because past medical conditions might influence current treatment decisions—eg, non-steroidal anti-inflammatory drugs for musculoskeletal disorders should be used with caution in patients with a history of gastrointestinal disorders. This approach is also crucial to understanding the mechanisms leading to multimorbidity because chains of disease causation might be separated in time—eg, acute events, such as myocardial infarction, arise from antecedent risk factors, such as hypertension, and can lead to later heart failure. Correspondingly, we defined comorbidity as the accumulation of additional conditions to an index condition (a specific condition under consideration) over an individual's lifetime.

Previous large-scale studies have used network analysis to examine the association between diseases.[4–7] However, many used hospital data alone,[4–6] which could overlook common conditions mainly diagnosed and treated in primary care, such as hypertension, mild-to-moderate depression, and soft-tissue disorders. Most studies have excluded children and young people, despite evidence that multimorbidity can affect younger people, patterns are likely to be different than in older age groups,[8,9] and conditions such as cardiovascular disease can be a late complication of childhood diseases, such as malignancy or its treatment. Also, few studies to date have considered associations by ethnicity or race, despite differences in environmental, biological, and socioeconomic risk factors between ethnic groups that could result in different patterns of multimorbidity.[2]

We addressed these gaps by analysing a large representative population dataset in England, UK, combining primary care and hospital admission records for 308 health conditions with expert-curated case definitions.[10] We extended our previous work mapping the course of human health for the most common single conditions in each decade of life[10] to multiple conditions by characterising multimorbidity and comorbidity for the whole population

and subgroups stratified by ethnicity, sex, and age in two distinct but complementary ways: identifying the most frequent combinations of conditions and using network analysis to discover non-random associations between conditions. We developed online interactive tools to allow others to visualise these associations in detail.

Methods

Study design and population

In this population-based study, we used primary care electronic health records (EHRs) from the Clinical Practice Research Datalink (CPRD) linked with the Hospital Episode Statistics admitted patient care dataset in England (appendix p 3). We included individuals in our study if they were older than 1 year, had been registered for at least 1 year in a participating general practice during the study period (from April 1, 2010, to March 31, 2015), and their practice and individual records met research data standards set by the CPRD. Follow-up was censored at the earliest event of death, de-registration from the practice, the last data collection for the practice, or March 31, 2015 (appendix p 3). We grouped self-reported race or ethnicity (hereafter referred to as ethnicity) into five categories: Black, Mixed, other, south Asian, and White. Individuals with missing ethnicity or with multiple conflicting ethnicity codes were classified as unknown. Results stratified by ethnic group are reported for Black, south Asian, and White subpopulations only, because Mixed and other subgroups were too heterogeneous for meaningful interpretations to be drawn.

The study was approved by the UK Medicines and Healthcare products Regulatory Agency Independent Scientific Advisory Committee [16_022]. Primary care practices provide consent for CPRD to collect deidentified data from their practice. Individual patients can opt out of sharing their data for research and CPRD does not collect data for these patients.

Health condition selection and case definition

We selected 308 physical and mental health conditions, as previously described (appendix pp 3–4, 8–12).[10] Phenotyping algorithms for these conditions are available online.[10,11] Individuals were considered to have ever been diagnosed with a particular condition if their records fulfilled the criteria in the algorithm for that condition before or during the study period.

Multimorbidity and comorbidity by frequency

We developed online tools to examine multimorbidity and comorbidity by frequency: the Multimorbidity Frequency App (MFA), which tabulates the 50 most prevalent triads (three conditions co-occurring in one individual), and the Comorbidity Frequency App (CFA), which shows the most prevalent comorbidities for an index condition (appendix p 13). These tools were developed using the shiny package in R (version 4.1.1).

We investigated the types of comorbidities that were typically excluded from heart failure trials. For this, we obtained proportions of heart failure trials that excluded patients with specific comorbidities from Boyd and colleagues' survey of trials of chronic diseases from Cochrane Reviews.[12]

Analysis of multimorbidity and comorbidity by non-random association

We used partial correlation to quantify the strength of association between two health conditions (appendix pp 4–5). Partial correlations greater than zero indicated that conditions were more likely to have ever occurred in the same individual than would be expected by chance. Higher partial correlations denoted stronger associations.

We did a network analysis in which networks comprising nodes (representing conditions) connected by edges were used to visualise the non-random associations between health conditions. Full details of this analysis are in the appendix (p 5).

We developed the Multimorbidity Network App (MNA) and Comorbidity Network App (CNA; appendix p 13). Multimorbidity networks consisted of all the connections between the 308 conditions for selected ϕ values (the specified partial correlation threshold for displaying the edges between two conditions; appendix p 5). The MNA allows users to find a balance between the degree of information conveyed and the clarity of visualisation displayed in the networks (appendix p 5). Comorbidity networks display conditions that were more likely to co-occur with an index condition than expected by chance for specified ϕ thresholds. The CNA allows non-random comorbidities to be interactively interrogated for any of the 308 index conditions. We did these analyses using the igraph and shiny packages in R (version 4.1.1).

We defined a multimorbidity coefficient as the sum of the positive partial correlations connecting a health condition to all the other health conditions in the study (appendix pp 5–6). This coefficient denotes how connected one condition is to all other conditions. The multimorbidity coefficient can also be interpreted as the number of connections (edges) with other conditions multiplied by the average positive partial correlation per connection (appendix pp 5–6).

For the **phenotyping algorithms** see https://www.caliberresearch.org/portal/phenotypes/chronological-map

Role of the funding source

The funders had no role in study design, data collection, data analysis, data interpretation, writing of the report, or the decision to submit the paper for publication.

Results

For the period April 1, 2010, to March 31, 2015, we derived multimorbidity and comorbidity patterns from 3 872 451 individuals, of whom 1 955 700 (50·5%) were female, 1 916 751 (49·5%) were male, 2 666 234 (68·9%) were White, 155 435 (4·0%) were south Asian, and 98 815 (2·6%) were Black (table).

	Population (N=3 872 451)	Age, years	Number of diagnosed conditions	Number of individuals with two or more diagnosed conditions (n=2 762 028)*
Sex				
Female	1 955 700 (50·5%)	38·76 (24·17)	5·91 (6·09)	1 468 214 (75·1%)
Male	1 916 751 (49·5%)	37·11 (22·97)	4·79 (5·60)	1 293 814 (67·5%)
Age groups, males and females (years)				
0–9	542 337 (14·0%)	3·15 (3·56)	1·84 (2·16)	239 488 (44·2%)
10–19	433 169 (11·2%)	14·62 (2·89)	2·41 (2·36)	248 369 (57·3%)
20–29	546 371 (14·1%)	24·64 (2·85)	2·81 (2·87)	329 294 (60·3%)
30–39	546 596 (14·1%)	34·48 (2·92)	3·48 (3·52)	359 717 (65·8%)
40–49	558 963 (14·4%)	44·44 (2·85)	4·98 (4·49)	432 873 (77·4%)
50–59	443 489 (11·5%)	54·28 (2·88)	6·91 (5·43)	384 222 (86·6%)
60–69	379 796 (9·8%)	64·15 (2·77)	9·37 (6·31)	355 629 (93·6%)
70–79	244 823 (6·3%)	74·22 (2·85)	12·73 (7·37)	238 240 (97·3%)
≥80	176 907 (4·6%)	85·47 (4·51)	15·48 (7·59)	174 196 (98·5%)
Race or ethnicity				
Black	98 815 (2·6%)	28·95 (19·88)	3·63 (4·42)	59 339 (60·1%)
South Asian	155 435 (4·0%)	28·98 (19·77)	3·98 (5·02)	93 617 (60·2%)
White	2 666 234 (68·9%)	39·98 (24·76)	6·45 (6·35)	2 097 536 (78·7%)
Mixed	33 673 (0·9%)	18·25(18·28)	2·72 (3·41)	17 835 (53·0%)
Other	58 019 (1·5%)	30·81 (19·54)	3·06 (4·09)	30 175 (52·0%)
Unknown	86 0275 (22·2%)	35·52 (19·75)	2·65 (3·11)	46 3425 (53·9%)
Overall	..	37·94 (23·60)	5·35 (5·88)	..

Data are n (%) or mean (SD). *Proportions calculated using the corresponding subgroup number in "Population" column as the denominator.

Table: **Characteristics of the study population**

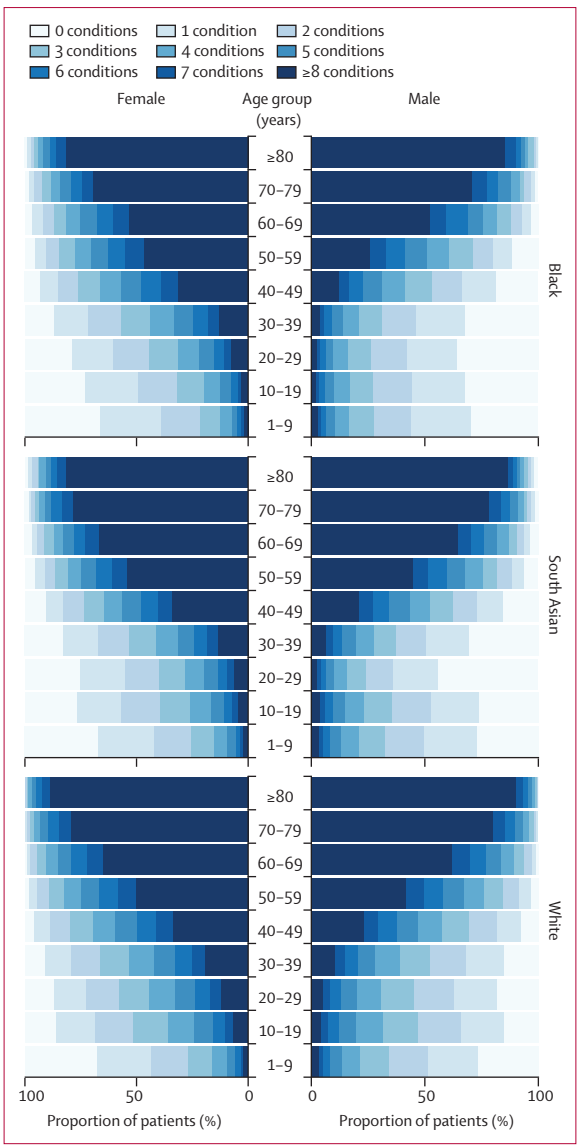


Figure 1: **Number of conditions ever recorded per individual (from 308 health conditions), stratified by age, sex, and ethnicity**

The number of diagnosed conditions per individual increases with age, with contrasting trajectories in females and males (figure 1; appendix pp 14–15, 96). 132 506 (47·8%) of 277 158 boys aged 1–9 years had been diagnosed with at least two conditions compared with 106 982 (40·3%) of 265 179 girls in the same age group. Among individuals aged 10 years and older, more women and girls than men and boys had two or more conditions (1 361 232 [80·5%] of 1 690 521 *vs* 1 161 308 [70·8%] of 1 639 593; figure 1). Among individuals aged 80 years and older, 111 157 (99%) of 112 673 women and 63 039 (98%) of 64 234 men had been diagnosed with at least two conditions during their lifetime. 2 097 536 (78·7%) of 2 666 234 White individuals had been recorded as having two or more conditions compared with 59 339 (60·1%) of 98 815 Black and 93 617 (60·2%) of 155 435 south Asian individuals (table). Higher recorded multimorbidity levels were observed in White individuals than in Black or south Asian individuals in each stratified subgroup by age and sex (figure 1).

The MFA can be used to view the 50 most common triads of conditions for selected subgroups (appendix pp 16–45). Conditions in the top triads corresponded with the most prevalent conditions in each subgroup (appendix p 7).

The CFA can be used to view the most common comorbidities for an index condition.

For the **Multimorbidity Frequency App (MFA)** see https://multimorbidity.caliberresearch.org/MFA

For the **Comorbidity Frequency App (CFA)** see https://multimorbidity.caliberresearch.org/CFA

Depression frequently co-occurred with anxiety, migraine, obesity, asthma, allergic rhinitis, dermatitis, deafness, soft-tissue disorders, and gastrointestinal disorders across all subgroups (appendix pp 103–10). Individuals with depression aged 20–29 years also frequently has a co-occurrence of acne (19 090 [22·9%] of 83 212 individuals with depression), menorrhagia (9870 [11·9%]), dysmenorrhoea (8542 [10·3%]), and substance misuse (7210 [8·7%]). Common comorbidities in individuals with depression aged 50–59 years were hypertension (35 018 [31·6%] of 110 830), increased LDL cholesterol (34 441 [31·1%]), osteoarthritis (24 464 [22·1%]), type 2 diabetes (12 213 [11·0%]), and chronic obstructive pulmonary disease (COPD; 7849 [7·1%]). Common comorbidities in individuals with depression

aged 80 years and older were cataracts (21 071 [55·6%] of 37 928), dementia (13 275 [35·0%]), chronic kidney disease (13 071 [34·5%]), atrial fibrillation (12 569 [33·1%]), heart failure (9815 [25·9%]), osteoporosis (10 235 [27·0%]), and myocardial infarction (6736 [17·8%]; appendix pp 103–07). Additional notable comorbid conditions in women and girls with depression were urinary incontinence (48 341 [11·8%] of 410 081), thyroid disease (48 073 [11·7%]), and dysmenorrhoea (41 228 [10·0%]), while additional notable comorbid conditions in men and boys with depression were erectile dysfunction (38 433 [16·7%] of 229 734), abdominal hernia (33 729 [14·7%]), benign prostatic hyperplasia (22 721 [9·9%]), and substance misuse (19 990 [8·7%]; appendix p 108). Common comorbidities for depression were similar across the three ethnic subgroups (appendix pp 109–10).

Clinical trials for heart failure frequently exclude individuals with coronary heart disease, valvular disease, arrhythmias, lung diseases, musculoskeletal diseases, hypertension, COPD, renal insufficiency, type 2 diabetes, and peripheral vascular disease.[12] In our analysis, we found that these conditions were very common in patients with heart failure, with 11·7% (multiple valve disorder) to 81·6% (hypertension) of patients living with these comorbidities (figure 2).

Multimorbidity network plots showing the strongest non-random associations (appendix p 7) for different ethnic groups showed that spinal fractures (collapsed vertebra) were most strongly associated with malignancy in Black individuals, but with osteoporosis in White individuals (appendix pp 124, 126). Hypertension was most strongly associated with kidney disorders for individuals aged 20–29 years, but with dyslipidaemia, obesity, and diabetes in those aged 40 years and older (appendix pp 113–26; these multimorbidity networks can be visualised with greater acuity using the MNA).

High multimorbidity coefficients were seen across ethnic subgroups for heart failure (Black individuals: 1·94 [IQR of multimorbidity coefficients across all conditions 0·42–0·99]; south Asian individuals: 2·04 [0·41–0·96]; and White individuals: 2·01 [0·39–1·00]), osteoarthritis (Black individuals: 1·19; south Asian individuals 1·54; and White individuals: 1·76), and liver cirrhosis (Black individuals: 1·48; south Asian individuals: 1·39; and White individuals: 1·44; figure 3; appendix pp 56–64).

However, slightly higher multimorbidity coefficients were seen for Black individuals versus south Asian or White individuals for leukaemia (Black individuals: 1·08; south Asian individuals: 0·73; and White individuals: 0·70) and lupus erythematosus (Black individuals: 1·01; south Asian individuals: 0·73; and White individuals 0·52; figure 3; appendix pp 56–64).

Higher multimorbidity coefficients were seen for White individuals than for Black and south Asian individuals for irritable bowel syndrome (White individuals: 1·23; Black individuals: 0·61; and south Asian individuals: 0·71) and asthma (White individuals: 1·03; Black individuals 0·70; and south Asian individuals: 0·75; figure 3; appendix pp 56–64).

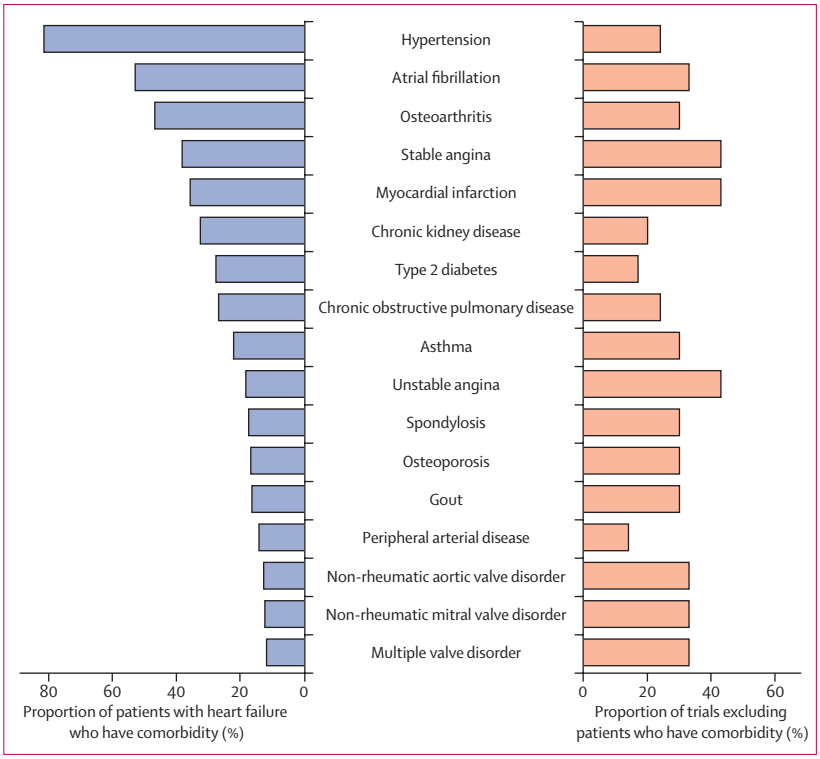


Figure 2: **Proportion of patients with heart failure with common comorbidities *vs* proportion of heart failure trials that exclude patients with these comorbidities**
Estimates derived from Boyd and colleagues.[12]

Women and girls had higher multimorbidity coefficients than did men and boys for rheumatoid arthritis (women and girls 0·96 [IQR of multimorbidity coefficients across all conditions 0·39–0·95] *vs* men and boys 0·76 [0·36–1·00]) and migraine (0·94 *vs* 0·56), whereas men and boys had higher coefficients than women and girls for abdominal hernia (men and boys: 0·98 *vs* women and girls 0·60) and actinic keratosis (1·03 *vs* 0·75; appendix pp 50–55).

Different conditions had high multimorbidity coefficients in different age groups (appendix pp 65–91). High coefficients for leukaemia (1·56 [IQR of multimorbidity coefficients across all conditions 0·23–0·90]) and brain cancer (1·39 [0·23–0·90]) were seen in individuals aged 1–9 years. High coefficients were seen for menorrhagia (1·74 [0·32–0·87]) and obesity (1·41 [0·32–0·87]) in individuals aged 40–49 years. And high coefficients were seen for benign prostatic hyperplasia (1·66 [0·21–0·69]) and diverticular disease (1·15 [0·21–0·69]) in individuals aged 80 years and older.

The number of connections and average partial correlation per connection for the top 100 multimorbidity

coefficients in each subgroup are in the appendix (pp 50–91).

The CNA can be used to explore comorbidity networks for an index disease.

To illustrate the similarities and differences in comorbidity patterns by ethnicity, sex, and age, we compared networks for the following exemplar conditions: primary breast cancer in Black, south Asian, and White individuals (both men and women); asthma in female and male individuals; and bipolar affective disorder in younger and older age groups (figure 4; appendix p 7).

Breast cancer in all three ethnic groups was associated with secondary lymph node cancer (partial correlation

For the **Comorbidity Network App (CNA)** see https://multimorbidity.caliberresearch.org/CNA

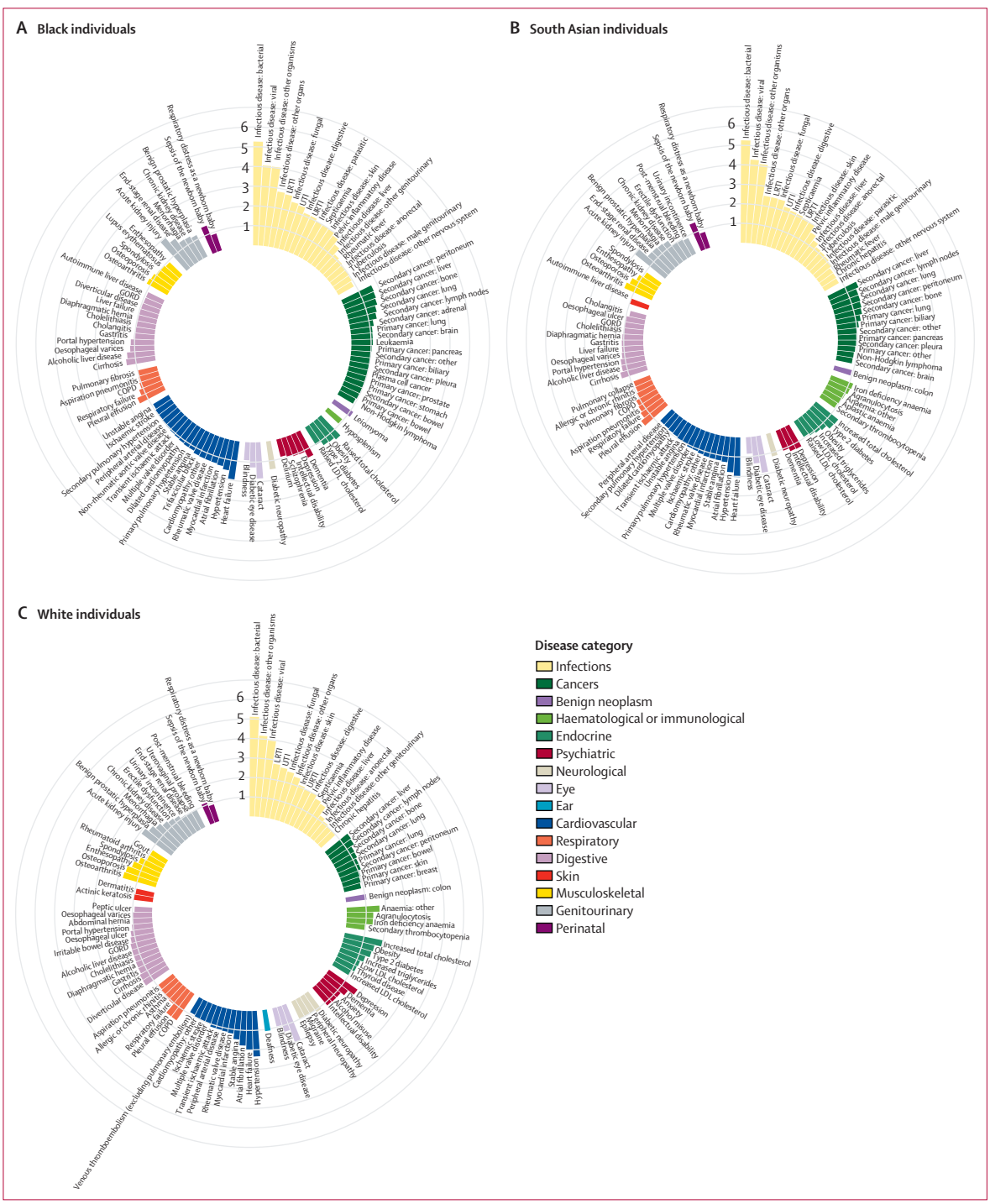
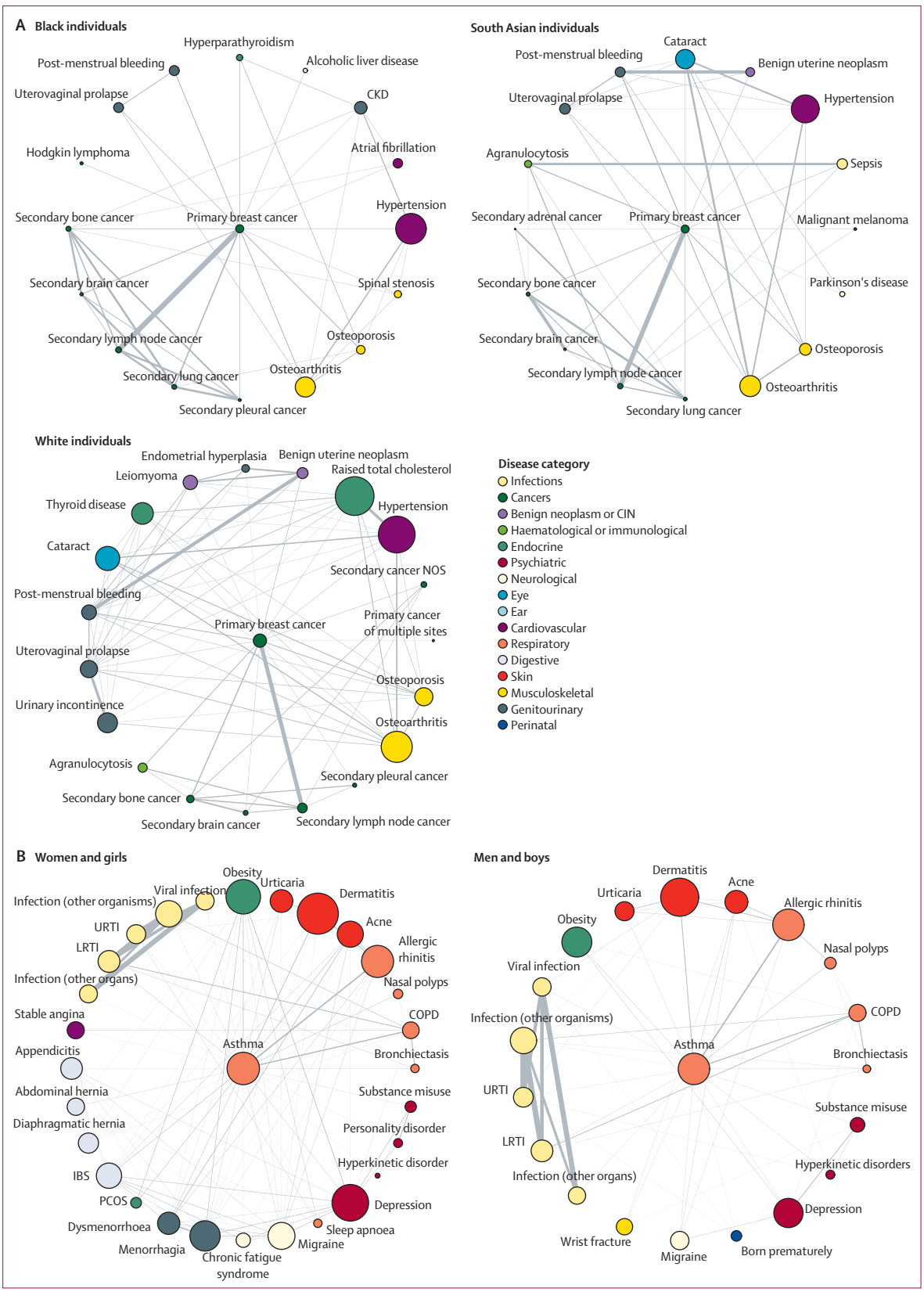


Figure 3: **Multimorbidity coefficient of the top 100 conditions for Black (A), south Asian (B), and White individuals (C)**

Each bar shows the multimorbidity coefficient in each population, shown on a scale of 0 to 6, with white bands indicating a coefficient of 1, 2, 3, 4, 5, or 6, as indicated on the scale. The multimorbidity coefficient is the sum of the positive partial correlations of a condition with all 307 other conditions in the subgroups. It quantifies the strength of a condition's association with other conditions in the study. A high coefficient indicates a condition that might share common risk factors with many other conditions, while a low coefficient suggests a condition that is unlikely to share disease pathways with other conditions. COPD=chronic obstructive pulmonary disease. GORD=gastro-oesophageal reflux disease. LRTI=lower respiratory tract infection. URTI=upper respiratory tract infection. UTI=urinary tract infection.



(Figure 4 continues on next page)

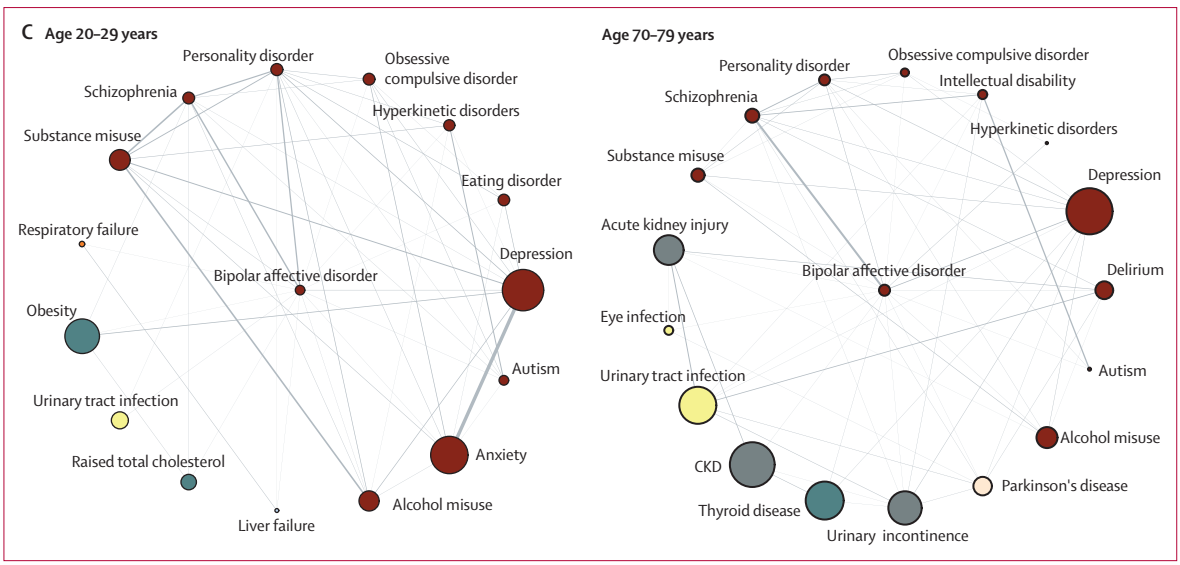


Figure 4: **Comparison of comorbidities with partial correlations >0·01 for primary breast cancer by ethnicity (A), asthma by sex (B), and bipolar affective disorder by age group (C)**

The size of a node is proportional to the period prevalence of the condition from April 1, 2010, to March 31, 2015, for each subgroup. The width of the edge between two nodes is proportional to the partial correlation between the two conditions represented by the nodes. CIN=cervical intraepithelial neoplasia. CKD=chronic kidney disease. COPD=chronic obstructive pulmonary disease. IBS=irritable bowel syndrome. LRTI=lower respiratory tract infection. NOS=not otherwise specified. PCOS=polycystic ovary syndrome. URTI=upper respiratory tract infection.

0·321 [IQR of partial correlations of all 308 comorbidities 0·0009–0·0051] for Black individuals; 0·288 [0·0009–0·0047] for south Asian individuals; 0·242 [0·0009–0·0044] for White individuals), secondary brain cancer (0·0516 for Black individuals; 0·0334 for south Asian individuals; and 0·0178 for White individuals), secondary bone cancer (0·0351 for Black individuals; 0·0411 for south Asian individuals; 0·0658 for White individuals), post-menopausal bleeding (0·0180 for Black individuals; 0·0495 for south Asian individuals; 0·0495 for White individuals), osteoporosis (0·0472 for Black individuals; 0·0491 for south Asian individuals; and 0·0477 for White individuals), hypertension (0·0221 for Black individuals; 0·0182 for south Asian individuals; 0·0226 for White individuals), and uterovaginal prolapse (0·0236 for Black individuals; 0·0185 for south Asian individuals; 0·0206 for White individuals). In Black individuals versus south Asian and White individuals, breast cancer was more likely to co-occur with secondary pleural cancer (0·0332 for Black individuals; <0·01 for south Asian and White individuals), hyperparathyroidism (0·0345 for Black individuals; <0·01 for south Asian and White individuals), Hodgkin lymphoma (0·0260 for Black individuals; <0·01 for south Asian and White individuals), alcoholic liver disease (0·0257 for Black individuals; <0·01 for south Asian and White individuals), and atrial fibrillation (0·0186 for Black individuals; <0·01 for south Asian and White individuals). South Asian individuals with breast cancer were more likely than Black or White individuals to have sepsis (0·0260 for south Asian individuals; <0·01 for Black and White individuals), malignant melanoma (0·0181 for south Asian individuals; <0·01 for Black and White individuals), and Parkinson's disease (0·0131 for south Asian individuals; <0·01 for Black and White individuals). White individuals with breast cancer were more likely than Black and south Asian individuals to have increased total cholesterol (0·0193 for White individuals; <0·01 for Black and south Asian individuals), thyroid disease (0·0165 for white individuals; <0·01 for Black and south Asian individuals), and leiomyoma (0·0145 for White individuals; <0·01 for Black and south Asian individuals; figure 4A).

Asthma had the following common comorbid conditions in both sexes: allergic rhinitis (partial correlation 0·158 [IQR of partial correlations of all 308 comorbidities 0·0002–0·0047] for women and girls *vs* 0·167 [0·0007–0·0056] for men and boys), COPD (0·132 *vs* 0·119), dermatitis (0·0653 *vs* 0·0893), obesity (0·0607 *vs* 0·0319), and lower respiratory tract infections (0·0310 *vs* 0·0239). Additionally, women and girls with asthma were more likely than men and boys to be diagnosed with irritable bowel syndrome (0·0170 for women and girls *vs* <0·01 for men and boys) and appendicitis (0·0148 *vs* <0·01), whereas men and boys with asthma were more likely than women and girls to have been diagnosed with wrist fracture (0·0194 for men and boys *vs* <0·01 for women and girls) and to have been born prematurely (0·0121 *vs* <0·01; figure 4B).

Bipolar affective disorder in individuals aged 20–29 years and those aged 70–79 years was more likely to occur with schizophrenia (partial correlation 0·0135 [IQR of partial correlations of all 308 comorbidities 0·001–0·0049] for individuals aged 20–29 years *vs* 0·195 [0·0019–0·0098] for individuals aged 70–79 years), personality disorder (0·122

vs 0·0409), and depression (0·0366 *vs* 0·0673). Individuals with bipolar affective disorder aged 20–29 years were also more likely than those aged 70–79 years to have increased total cholesterol (0·0123 for individuals aged 20–29 years *vs* <0·01 for individuals aged 70–79 years), liver failure (0·0115 *vs* <0·01), and obesity (0·0103 *vs* <0·01), whereas those aged 70–79 years with bipolar affective disorder were more likely than those aged 20–29 years to have chronic kidney disease (0·0137 for individuals aged 70–79 years *vs* <0·01 for individuals aged 20–29 years), acute kidney injury (0·0124 *vs* <0·01), and thyroid disease (0·0115 *vs* <0·01; figure 4C).

Discussion

Using data from almost 4 million patients in the English National Health Service, we identified differences in multimorbidity and comorbidity patterns by age, sex, and ethnicity, with implications for clinical practice, service provision, clinical trial design, drug development, and research into shared disease pathways. This study constitutes a major extension of our previous work mapping single diseases across the human life course and is, to our knowledge, the first in-depth description of multimorbidity and comorbidity patterns for hundreds of curated disease phenotypes across all organ systems in a large representative study population in England from the dual perspectives of frequency and non-random disease association. We also provide online tools to enable others to interrogate these patterns in more detail.

The proportion of individuals with two or more documented conditions was higher in White individuals than in Black and south Asian individuals even after stratification by age. Comparison with previous reports is complicated by differences in sample characteristics (eg, differences in the geographical location, racial or ethnic, and socioeconomic constituents of the sample population, and whether the sample population was obtained from insurance databases, primary or secondary care health records, or research cohorts). However, another study using data from England found White individuals had greater multimorbidity than people in other ethnic groups,[13] whereas two US studies found that a higher proportion of Black individuals had multimorbidity than did White and Hispanic or Asian individuals.[14,15] Further investigation should ascertain whether our findings are due to more frequent consultations, preferential access to health care, bias in recording of diagnostic codes, or truly higher multimorbidity for White individuals than individuals of other ethnicities.

The multimorbidity coefficient summarises the strength of association between one condition and all other conditions in the multimorbidity network, thus indicating which conditions are associated with the greatest burden of non-random multimorbidity in the different subgroups. Patients aged 1–9 years with cancer, patients aged 40–49 years with menorrhagia and obesity, and patients aged 80 years and older with benign prostatic hyperplasia and diverticulosis were more likely to have other conditions, and hence have increased health needs. Differences in multimorbidity coefficients between the sexes, with women and girls having higher coefficients for rheumatoid arthritis and migraine, and men and boys having higher coefficients for abdominal hernia and actinic keratosis, could suggest differential exposure to risk factors or different disease pathways. Different disease mechanisms might also explain why Black individuals had higher coefficients for leukaemia and lupus erythematosus than individuals in other ethnic groups.

Bipolar affective disorder was associated with increased total cholesterol, liver failure, and obesity in younger patients (aged 20–29 years), whereas kidney disorders and thyroid disease were common in older patients with this condition (aged 70–79 years). Metabolic changes are a well known side-effect of antipsychotic drugs, which were authorised in 2000 by the US Food and Drug Administration to treat bipolar affective disorder, whereas kidney disorders and thyroid disease are recognised adverse effects of lithium, which was used more commonly before this date.[16,17] Therefore, these comorbidities might be induced by treatment. Another possibility is that individuals with metabolic disorders had shorter lifespans, so older individuals without metabolic comorbidities presented a survival bias. Further investigations are required to establish why the two age groups had different comorbidity patterns—one of the evidence gaps identified in the 2018 Academy of Medical Sciences report on multimorbidity.[2]

We found different comorbidities to be associated with asthma between the sexes, consistent with previous studies.[18] Asthma is more common and severe in women and girls after puberty, suggesting that sex hormones have a role in the pathophysiology of asthma subtypes, as reflected in the different age profiles of male and female individuals who develop this condition.[10,19]

Osteoporosis and breast cancer were strongly associated with each other in all ethnic groups, probably because of the use of oestrogen-suppressing therapy for breast cancer.[20] However, Black individuals with breast cancer also had a strong non-random association with atrial fibrillation—an association that has been reported previously.[21] Further studies are needed to establish whether this association with atrial fibrillation in Black individuals is due to more aggressive breast cancer or delayed diagnosis. Additionally, we found that Parkinson's disease is associated with breast cancer in south Asian individuals, the reason for which is unclear, and previous studies into this link have been inconclusive.[22] Future studies need to investigate whether these conditions share disease mechanisms or risk factors, whether treatment for one condition could lead to the other, and whether these factors are specific to south Asian individuals.

The UK National Institute for Health and Care Excellence has issued guidelines for the management of depression in adults with a chronic physical health

problem.[23] However, these guidelines do not discuss how to manage patients with depression with common comorbidities, such as obesity, dyslipidaemia, hypertension, atopic conditions, soft-tissue disorders, and osteoarthritis. One synergistic therapeutic approach could be to invest in physical activity programmes that can simultaneously tackle depression and obesity, dyslipidaemia, and hypertension, and potentially prevent conditions resulting in chronic pain, such as soft-tissue disorders and osteoarthritis.

Heart failure is a condition of old age and thereby associated with greater multimorbidity. We found that a large proportion of heart failure trials exclude patients with common comorbidities. Similar data have been found for COPD, diabetes, stroke, cancers, depression, dementia, arthritis, and atrial fibrillation,[12,24–26] thus limiting the real-world applicability of clinical trials and related guidelines.

Our study benefits from use of a large generalisable EHR dataset that reflects the clinical reality and diversity of individuals in England and access to 308 curated phenotypes covering a wide spectrum of health conditions. Stratifying by ethnicity, sex, and age allowed us to investigate how comorbidities differed between subgroups for the same condition.[2]

In a reasoned departure from the conventional view of multimorbidity as the coexistence of two or more chronic conditions, we analysed patients' cumulative medical histories to detect conditions that, although separated in time, might share pathological mechanisms. For example, the association of premature birth with asthma might reflect a developmental cause in male individuals.[27] We did not model disease progression in this study; therefore, the sequence of disorders cannot be determined from our findings. However, comorbidity networks for stratified age groups enabled us to observe different comorbidities in different age groups. Analysing comorbidities by age also allowed us to identify comorbid clusters in younger patients—another need highlighted in the 2018 Academy of Medical Sciences report.[2] We also found networks in which treatments for one condition could possibly lead to multiple other conditions depending on historical prescribing patterns, such as in bipolar affective disorder. Therefore, our study has the potential to generate many hypotheses such as these for future investigations.

Direct comparison of the prevalence and patterns of multimorbidity across studies is complicated by considerable heterogeneity in methods and definitions.[2,28] As advocated for by Whitty and Watt,[29] we took a hypothesis-free approach, clearly reporting our case definitions for a wide range of health conditions so that our findings can be interpreted within the context of our stated aims: identifying frequently occurring disease combinations that could inform clinical management and non-random disease associations that could indicate shared disease mechanisms.

Our study has several limitations. Our results depend on the accuracy of EHR data, as previously discussed.[10] Briefly, some conditions might be under-represented in EHRs if they are asymptomatic, treated in specialist clinics not linked to EHRs, or can be self-managed; delays might occur between disease onset and documentation of a diagnosis, and these delays might be greater in some subgroups than in others due to different health-seeking behaviours and preconceived clinician biases; some conditions might be diagnosed while investigating other conditions; screening or public health programmes might bias the diagnosis of some conditions to specific demographic groups; or conditions could have been incorrectly coded or misdiagnosed. Therefore, these caveats should be considered when interpreting our results. However, by identifying these associations in the EHRs, our networks can inform researchers about possible inconsistencies in the recorded data.

Our dual-axis classification for infectious conditions by causal organism and organ system could lead to infections being more highly represented in disease triads, multimorbidity networks, and multimorbidity coefficients than other conditions. This fact should be considered when interpreting our findings.

We did not have access to individual socioeconomic information in our dataset. However, numerous studies have already shown that socioeconomic status affects multimorbidity patterns.[30,31] Our analysis addresses the evidence gap in multimorbidity patterns for different ethnic groups and younger individuals. We also did not analyse the association between mortality and multimorbidity because it was beyond the scope of this study.

The COVID-19 pandemic and subsequent disruption to health services might have led to new multimorbidity patterns since 2020. Future studies need to examine the impact of the pandemic on multimorbidity patterns.

Biomedical research typically focuses on single conditions with little attention to overlapping pathological mechanisms. Non-random associations that we identified in this study or that might be identified by users of our online tools could provide the foundation for investigating shared disease mechanisms and risk factors. Overlaps in genetic loci for multiple phenotypes in genetic association studies suggest that any given therapeutic target could influence the risk of multiple conditions.[32] Phenotypes with associations to the same drug target genes that coincide with non-random disease associations from our study could lead to drug repurposing opportunities or the development of drugs that can treat multiple conditions concurrently. Similarly, researchers could investigate non-pharmacological interventions of risk factors that affect non-random multimorbidities. Additionally, research establishing which comorbidities are caused by the adverse effects of therapeutic drugs or procedures could give patients and clinicians much needed information to weigh the

benefits and harms of different treatment options. Awareness of an increased risk of specific comorbidities could alert clinicians to screen for these comorbidities and precipitate research into interventions that might prevent these conditions.

Clinical trial design and guideline development need to change to reflect the reality of the increasing multimorbidity of our current and future populations. Single-disease guidelines based on clinical trials that exclude patients with comorbidities often make contradictory and impractical recommendations for patients with multimorbidities.[3] Instead, clinical guidelines should advise on treatment synergies, ways to minimise adverse interactions between drugs and diseases, conflicting management strategies, and polypharmacy for patients with common comorbidities. Health-care services and staff training can be organised around common multimorbidities, thus improving care for the most susceptible patients. Our interactive tool can help researchers identify common comorbidities for multiple conditions.

Minority ethnic groups and younger patients have typically been overlooked in multimorbidity research. By including these subgroups and stratifying our analyses according to ethnicity, sex, and age, we hope to address the paucity of evidence in this area. By enabling researchers to explore these differences, our tools provide the groundwork for future research.

Contributors

VK, ADH, and HH conceived the project. VK designed the study, did the analysis, developed the online tools, interpreted the data, and wrote the original draft. ADH and HH supervised the work. SD constructed the website for the open access resources. The terms of University College London (UCL)'s multi-study licence and Independent Scientific Advisory Committee (ISAC) project approval with CPRD are such that only individuals involved in directly managing or analysing the data are allowed to access the data, hence only VK, SD, AG-I, HH, and ADH had access to the raw data (furthermore VK, AG-I, and SD verified the underlying study data). However, all authors had access to detailed aggregated data and statistical outputs. SD and AG-I extracted the data and maintain the CALIBER platform. ADH obtained funding and led the project administration. All authors reviewed and interpreted the results, commented on the paper, contributed to revisions, and read and approved the final version. VK had final responsibility for the decision to submit for publication.

Declaration of interests

DN is the UK Kidney Association Director of Informatics Research based at the UK Renal Registry and is on the steering committee for two GlaxoSmithKline-funded studies looking at kidney function markers in sub-Saharan Africa. ICKW was a member of the ISAC of CPRD and has received funding from Amgen, Bristol-Myers Squibb, Pfizer, Janssen, Bayer, GSK, and Novartis to conduct pharmacoepidemiological research outside the submitted work. RM has received consulting fees from Amgen. ADH is a co-investigator on a grant from Pfizer to identify potential therapeutic targets for heart failure using human genomics. NC is remunerated for her membership of a data safety and monitoring committee of a trial sponsored by AstraZeneca. All other authors declare no competing interests.

Data sharing

Summary-level data are provided in the appendix (pp 14–93) and on our shiny apps interactive tools. Partial pairwise correlations for each stratified subgroup are available on the CALIBER platform. Code for the shiny apps tools is available on GitHub.

Acknowledgments

The study, VK, and AR are supported by the UK Research and Innovation (UKRI) Strategic Priority Fund Tackling multimorbidity at scale programme (grant number MR/V033867/1) delivered by the Medical Research Council and the National Institute for Health and Care Research (NIHR) in partnership with the Economic and Social Research Council and in collaboration with the Engineering and Physical Sciences Research Council. Additional support was provided by the Rosetrees Trust. ADH and HH are NIHR Senior Investigators. SD is supported by the British Heart Foundation (BHF) Data Science Centre (grant number SP/19/3/34678), the NIHR-UKRI CONVALESCENCE study, and the Longitudinal Health and Wellbeing COVID-19 National Core Study, which was established by the UK Chief Scientific Officer in October, 2020, and funded by UKRI (grant references MC_PC_20030 and MC_PC_20059). SD is supported by the BigData@Heart Consortium, funded by the Innovative Medicines Initiative-2 Joint Undertaking (under grant agreement number 116074). SD, HH, and ADH are funded by the NIHR UCL Hospitals Biomedical Research Centre and supported by the UCL BHF Research Accelerator (grant number AA/18/6/34223). Work at Health Data Research UK (award reference LOND1) is funded by the UK Medical Research Council, Engineering and Physical Sciences Research Council, Economic and Social Research Council, Department of Health and Social Care (England), Chief Scientist Office of the Scottish Government Health and Social Care Directorates, Health and Social Care Research and Development Division (Welsh Government), Public Health Agency (Northern Ireland), British Heart Foundation, and Wellcome Trust. RM is supported by a Sir Henry Wellcome Postdoctoral Fellowship from the Wellcome Trust (WT 201375/Z/16/Z). This study was carried out as part of the CALIBER programme. CALIBER, led from the UCL Institute of Health Informatics, is a research resource consisting of anonymised, coded variables extracted from linked electronic health records, methods and tools, specialised infrastructure, and training and support. This study is based in part on data from the CPRD obtained under licence from the UK Medicines and Healthcare products Regulatory Agency. The data are provided by patients and collected by the EnglishNational Health Service (NHS) as part of their care and support. The interpretation and conclusions contained in this study are those of the authors alone. Hospital Episode Statistics data were re-used with the permission of the Health and Social Care Information Centre. The Office of Population Censuses and Surveys Classification of Interventions and Procedures, codes, terms, and text are Crown copyright (2016), published by the Health and Social Care Information Centre, also known as NHS Digital, and licensed under the Open Government Licence available online.

For more on the **CALIBER programme** see https://www.ucl.ac.uk/health-informatics/caliber

For **Open Government licenses** see www.nationalarchives.gov.uk/doc/open-government-licence/open-government-licence.htm

References

1 WHO. Multimorbidity. Geneva: World Health Organization, 2016. http://apps.who.int/iris/handle/10665/252275 (accessed June 11, 2022).
2 The Academy of Medical Sciences. Multiple long-term conditions (multimorbidity): a priority for global health research. 2018. https://acmedsci.ac.uk/policy/policy-projects/multimorbidity (accessed June 11, 2022).
3 Guthrie B, Payne K, Alderson P, McMurdo MET, Mercer SW. Adapting clinical guidelines to take account of multimorbidity. *BMJ* 2012; **345:** e6341.
4 Jensen AB, Moseley PL, Oprea TI, et al. Temporal disease trajectories condensed from population-wide registry data covering 6.2 million patients. *Nat Commun* 2014; **5:** 4022.
5 Hidalgo CA, Blumm N, Barabási A-L, Christakis NA. A dynamic network approach for the study of human phenotypes. *PLOS Comput Biol* 2009; **5:** e1000353.
6 Siggaard T, Reguant R, Jørgensen IF, et al. Disease trajectory browser for exploring temporal, population-wide disease progression patterns in 7.2 million Danish patients. *Nat Commun* 2020; **11:** 4952.
7 Aguado A, Moratalla-Navarro F, López-Simarro F, Moreno V. MorbiNet: multimorbidity networks in adult general population. Analysis of type 2 diabetes mellitus comorbidity. *Sci Rep* 2020; **10:** 2416.
8 St Sauver JL, Boyd CM, Grossardt BR, et al. Risk of developing multimorbidity across all ages in an historical cohort study: differences by sex and ethnicity. *BMJ Open* 2015; **5:** e006413.

For **stratified subgroup correlations** see https://multimorbidity.caliberresearch.org/assets/pcor_csvfiles.tar.bz2

For **code for the shiny apps tools** see https://github.com/spiros/caliber-multimorbidity-shiny

9 Bobo WV, Yawn BP, St Sauver JL, Grossardt BR, Boyd CM, Rocca WA. Prevalence of combined somatic and mental health multimorbidity: patterns by age, sex, and race/ethnicity. *J Gerontol A Biol Sci Med Sci* 2016; **71:** 1483–91.

10 Kuan V, Denaxas S, Gonzalez-Izquierdo A, et al. A chronological map of 308 physical and mental health conditions from 4 million individuals in the English National Health Service. *Lancet Digit Health* 2019; **1:** e63–77.

11 Denaxas SC, George J, Herrett E, et al. Data resource profile: cardiovascular disease research using linked bespoke studies and electronic health records (CALIBER). *Int J Epidemiol* 2012; **41:** 1625–38.

12 Boyd CM, Vollenweider D, Puhan MA. Informing evidence-based decision-making for patients with comorbidity: availability of necessary information in clinical trials for chronic diseases. *PLoS One* 2012; **7:** e41601.

13 Mujica-Mota RE, Roberts M, Abel G, et al. Common patterns of morbidity and multi-morbidity and their impact on health-related quality of life: evidence from a national survey. *Qual Life Res* 2015; **24:** 909–18.

14 Glicksberg BS, Li L, Badgeley MA, et al. Comparative analyses of population-scale phenomic data in electronic medical records reveal race-specific disease networks. *Bioinformatics* 2016; **32:** i101–10.

15 St Sauver JL, Boyd CM, Grossardt BR, et al. Risk of developing multimorbidity across all ages in an historical cohort study: differences by sex and ethnicity. *BMJ Open* 2015; **5:** e006413.

16 Hayes J, Prah P, Nazareth I, et al. Prescribing trends in bipolar disorder: cohort study in the United Kingdom THIN primary care database 1995-2009. *PLoS One* 2011; **6:** e28725.

17 López-Muñoz F, Shen WW, D'Ocon P, Romero A, Álamo C. A history of the pharmacological treatment of bipolar disorder. *Int J Mol Sci* 2018; **19:** 2143.

18 Zein JG, Erzurum SC. Asthma is different in women. *Curr Allergy Asthma Rep* 2015; **15:** 28.

19 Tan DJ, Walters EH, Perret JL, et al. Age-of-asthma onset as a determinant of different asthma phenotypes in adults: a systematic review and meta-analysis of the literature. *Expert Rev Respir Med* 2015; **9:** 109–23.

20 Gralow JR, Biermann JS, Farooki A, et al. NCCN Task Force Report: bone health in cancer care. *J Natl Compr Canc Netw* 2013; **11** (suppl 3)**:** S1–50.

21 Guha A, Fradley MG, Dent SF, et al. Incidence, risk factors, and mortality of atrial fibrillation in breast cancer: a SEER-Medicare analysis. *Eur Heart J* 2022; **43:** 300–12.

22 Wang T. The link between Parkinson's disease and breast and prostate cancers: a meta-analysis. *Int J Neurosci* 2015; **125:** 895–903.

23 National Institute for Health and Care Excellence. Depression in adults with a chronic physical health problem: recognition and management (CG91). 2009. https://www.nice.org.uk/guidance/cg91 (accessed June 11, 2022).

24 He Z, Charness N, Bian J, Hogan WR. Assessing the comorbidity gap between clinical studies and prevalence in elderly patient populations. *IEEE EMBS Int Conf Biomed Health Inform* 2016; **2016:** 136–39.

25 Buffel du Vaure C, Dechartres A, Battin C, Ravaud P, Boutron I. Exclusion of patients with concomitant chronic conditions in ongoing randomised controlled trials targeting 10 common chronic conditions and registered at ClinicalTrials.gov: a systematic review of registration details. *BMJ Open* 2016; **6:** e012265.

26 Unger JM, Hershman DL, Fleury ME, Vaidya R. Association of patient comorbid conditions with cancer clinical trial participation. *JAMA Oncol* 2019; **5:** 326–33.

27 Goyal NK, Fiks AG, Lorch SA. Association of late-preterm birth with asthma in young children: practice-based study. *Pediatrics* 2011; **128:** e830–38.

28 Fortin M, Stewart M, Poitras ME, Almirall J, Maddocks H. A systematic review of prevalence studies on multimorbidity: toward a more uniform methodology. *Ann Fam Med* 2012; **10:** 142–51.

29 Whitty CJM, Watt FM. Map clusters of diseases to tackle multimorbidity. *Nature* 2020; **579:** 494–96.

30 Barnett K, Mercer SW, Norbury M, Watt G, Wyke S, Guthrie B. Epidemiology of multimorbidity and implications for health care, research, and medical education: a cross-sectional study. *Lancet* 2012; **380:** 37–43.

31 Charlton J, Rudisill C, Bhattarai N, Gulliford M. Impact of deprivation on occurrence, outcomes and health care costs of people with multiple morbidity. *J Health Serv Res Policy* 2013; **18:** 215–23.

32 Hingorani AD, Kuan V, Finan C, et al. Improving the odds of drug development success through human genomics: modelling study. *Sci Rep* 2019; **9:** 18911.